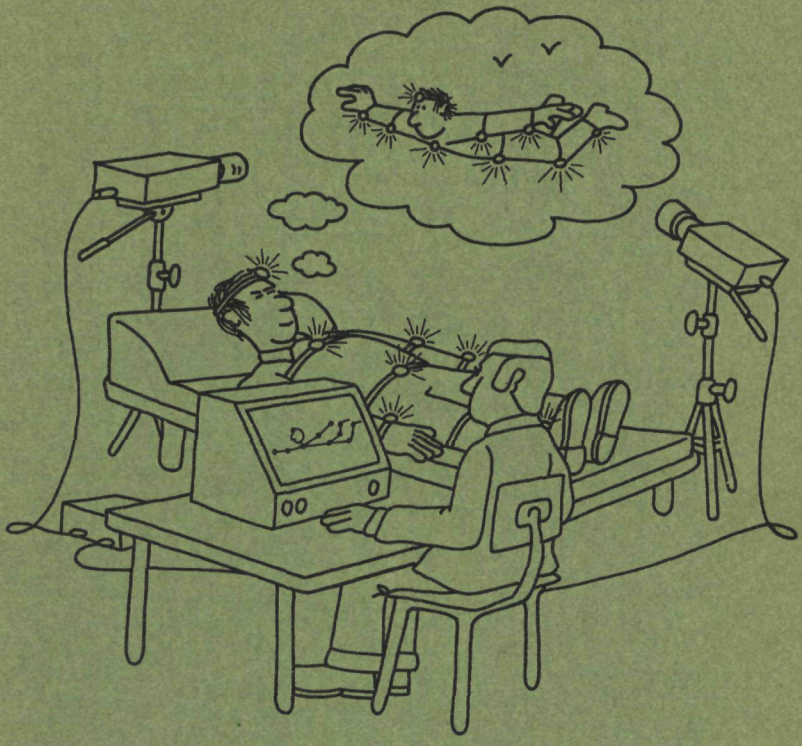


MEASUREMENT AND CONTROL OF HUMAN MOVEMENT

H.J. Woltring

An object is at rest when it occupies a space
equal to itself.
An arrow in flight occupies, at any given moment,
a space equal to itself.
Therefore, an arrow in flight is at rest.

Zeno of Elea's third syllogism against
the existence of motion.

PROMOTOR:
PROF. DR. IR. E.G.J. EIJKMAN

CO-PROMOTOR:
PROF. DR. F.L. STUMPERS

MEASUREMENT AND CONTROL OF HUMAN MOVEMENT

P R O E F S C H R I F T

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE
WISKUNDE EN NATUURWETENSCHAPPEN
AAN DE KATHOLIEKE UNIVERSITEIT TE NIJMEGEN, OP GEZAG VAN
DE RECTOR MAGNIFICUS, PROF. DR. A.J.H. VENDRIK,
VOLGENS BESLUIT VAN HET COLLEGE VAN DECANEN
IN HET OPENBAAR TE VERDEDIGEN
OP VRIJDAG 13 MEI 1977
DES NAMIDDAGS TE 2 UUR PRECIES

door

HERMAN JOZEF WOLTRING
geboren te Amsterdam

Druk: H. Peters & J. Haarsma
Nijmegen

Acknowledgements

I am indebted to the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) and to the Royal Society of London for a European Science Exchange Fellowship in 1971/1972 which enabled me to follow a postgraduate course in Experimental Psychology at the University of Sussex, England; in addition I would like to acknowledge support from Z.W.O. and Selcom A.B., Sweden, for a study voyage in 1976 to Canada and to the United States of America.

As regards the non-published parts of the present thesis, I would like to thank Mr. A. Fekkes for the cover illustration, Mr. R. Gras for photograph processing, Mrs. J.A. Thomassen-Baker for linguistical corrections, and Miss Jo Kiem Tioe for careful and patient typesetting during many evening hours. In addition, I am grateful to Mr. H. Peeters and Mr. J. Haarsma for many overtime hours in meeting the deadline for publication of the printed thesis.

I appreciate the permission granted by the copyright holders to reprint or preprint a number of journal papers: S. Karger A.G. of Basel for the Chapters 1.2, 2.1.1 and 2.1.2, the Institute of Electrical and Electronics Engineers of New York for Chapter 1.3, and Elsevier Scientific Publishing Company of Amsterdam for Chapter 2.2; moreover, I would like to thank the editors and printers of Photogrammetria for speeding up the reviewing and printing of the latter chapter.

Last but not least, I thank my wife for her support, and I apologize to her for the neglect which would seem to be an occupational risk of contemporary Ph.D. research.

aan Margriet
Chantal, Sjoerd
aan mijn ouders

Table of Contents

ERRATUM

p. 6, 12th line from below: for "calibratiom" read "calibration"

The pages 116 and 118 should be interchanged

p. 175, for "Galileo" read "Galilei"

Chapter I

Introduction and Data Acquisition Aspects

There are six sorts of movement: generation, destruction, increase, decrease, diminution, alteration, and change of place.

Aristotle, Categoriae 15^a14

1.1

Introduction

The study of human movement in all its spatio-temporal complexity, or *kinesiology*, has been somewhat of a Cinderella in a variety of disciplines including psychology and medicine. Most research in experimental psychology has been confined to simple movements along a line or in a plane where trajectories in tracking some reference signal are measured, usually of only one corporal point on hand or finger (see Van Assen, 1971, for a review). Interest is typically directed to underlying processes including stimulus analysis, response selection, memory aspects and prediction, and these may often be investigated by means of relatively simple movements in one or two dimensions. It was only in more applied contexts that movements used to be analysed in their full complexity, as with the industrial Time and Motion studies of the Gilbreths (1911, 1917). More recently, industrial automation and increased interest in the scientific fundamentals of sports and physical education have led to a shift to the latter field.

In medicine, Van Hussen (1973) has noted an increasing interest in human movement studies in order to improve prosthesis design; he further indicated that orthopaedical interest in kinesiology is insufficient, and that movement abilities have been measured up till now in a nonfunctional fashion, viz. on the operation table.

An important reason for such limited interest has been the lack of proper instrumentation and methodology in movement research. Current methods are predominantly based on photographic and cinematographic registration, followed by a tedious and errorprone datareduction procedure of digitizing film data if subsequent quantitive analysis is required, and the additional complexity of 3-dimensional (3-D) movement *reconstruction* from projective imagery has deterred many an investigator. Of course, many applications do not require highly accurate measurements; Vierordt (1881), for example, described

an interesting method based on spraying thin ink jets from the subject's body to long strips of paper lying on the ground and hung along the subject's walking path.

The tediousness of data processing has been greatly reduced because of contemporary computer developments, and recent advances in optoelectronics make it possible to bypass film registration; moreover, automatic identification of multiple landmarks on a moving subject is becoming practicable, either by pattern recognition techniques at the receiving end or by multiplexing active landmarks on the body in the time or frequency domains. The former approach is similar to those employed in, for instance, star identification in astronomy, particle trajectory reconstruction in high-energy physics, and aircraft recognition in radar navigation or fire-control; the latter method requires small light sources on the subject's body which are either switched on and off in turn (*time*-multiplexing) or modulated at individual frequencies (*frequency*-multiplexing) - a particularly interesting approach from a real-time point of view, albeit at the cost of wired markers on the subject's body.

Initially, the research leading to the present thesis was intended to involve limited work on instrumentation and methodology, in order to expand upon current research in human motor control from the point of view of experimental psychology. After a cursory definition of the type of research to be conducted, an investigation was started into instrumentation which would ultimately allow automated, full-bound analysis of human movement patterns in a real-time context (Chapter 1.2). This work resulted in an experimental system based on the *lateral photoeffect* (Wallmark, 1957) enabling time-multiplexed identification of infrared light emitting diodes (IR LEDs) which served as landmarks on the body. Since it appeared that a similar system was to become commercially available shortly, this development was terminated; in anticipation of this equipment, a formal mathematical analysis was conducted of the lateral photodetectors employed in both systems (Chapter 1.3), and of the software aspects of calibratiom and measurement in 3-D movement reconstruction.

It became increasingly clear that proper methodology in 3-D data processing on the *kinematic* level in kinesiology (trajectory estimation, derivative assessment) was insufficiently available; standard handbooks in biomechanics and kinesiology typically cover the equations of motion and methods to assess anthropometric parameters in order to calculate *kinetic* quantities (forces, torques, energy flows) during motion, but a unifying theory to assess the necessary kinematic quantities was found wanting. For this reason, an attempt has been made to fill the gap, the more so because suitable methods may be adopted from other disciplines; thus camera *calibration* and 3-D reconstruction from centrally projected imagery is the domain of *photogrammetry* which

has found its major application in civil engineering (surveying, cartography); signal differentiation from noisy data has been substantially developed in the telecommunications field and in control theory; and contemporary methods in *time domain filtering* since the work by Wiener, Kolmogorov, Kalman and Bucy has been extensively fostered by navigational interest in simultaneous estimation of 3-D trajectories and their time derivatives, based on direction and distance data obtained from radar or camera systems.

The papers in Chapter 2 contain an eclectic review of established facts from these parent disciplines, and new material on numerically efficient calibration procedures. From a practical point of view, calibration methods must be flexible and simple to use, and they must impose low computational demands, especially if frequent recalibration of camera configurations as may occur in field studies is required. Unlike the case in standard stereophotogrammetric equipment, there is the possibility of freely selecting camera *stations* and *attitudes*. This means that a greater depth accuracy and minimal risk of data loss due to shadowing effects may be simultaneously realized. For example, one might observe a suitably chosen distribution of calibration points prior to a series of actual measurements on human movement, and record these observations together for subsequent analysis. As regards the electronic methods currently available, such registration in the field is becoming increasingly feasible because of new coding techniques in tape-recording, i.e., PCM or *Pulse Code Modulation*. Therefore, it is not necessary to require on-line computer facilities as were actually available in most work reported in the following chapters; only if *real-time feedback* to subject or investigator is desired do such facilities become essential.

Some parts of the methodological work in Chapter 2 have been applied to the investigation of human motor control. In Chapter 3.1, a start has been made with research on how man is able to employ information which arrives during movement *execution*, for readjustment of his on-going "motor-programme", and this work will be generalized to more realistic, 3-D movement patterns in the future: certain hardware limitations in currently available computer facilities have held back research of this kind of real-time data processing.

Yet, in order to balance methodology and application in the present thesis, some pilot studies on a non real-time scale are presented in Chapter 3.2; here, the use of correlation functions shows aspects of posture preservation during one-legged and biped locomotion. Some tentative conclusions on the generality of a single "motor-programme" in such complex movements are presented; however, future research should be conducted on the higher level of kinetic analysis.

If applied work is typically preceded by a chapter on instruments and methods, a survey of applications following a methodological treatise would seem appropriate; Chapter 4 presents a somewhat speculative survey of history and trends in psychology and in related disciplines, following a similar section on instrumentation for continuous observation of human movement.

References

Gilbreth, F.B.: Motion Study (Van Nostrand, New York 1911).

Gilbreth, F.B. and Gilbreth, L.M.: Applied Motion Study (McMillan, New York 1917).

Van Assen, A.: Tracking Behaviour Tracked. Doct. Diss. University of Nijmegen, Lab. of Psychology (Nijmegen 1971).

Van Hussen, F.A.J.: Electrogoniometrie van de Knie - een Klinische Kinesiologische Studie. Doct. Diss., University of Nijmegen, Faculty of Medicine (Zutphen 1973).

Wallmark, J.T.: A new semiconductor photocell using lateral photo-effect. Proc. IRE *45*(1957) 474-483.

Biotelemetry
1 132-146 (1974)

The first creature of God, in the works of the days, was the light of the senses, the last was the light of reason.

Francis Bacon, Essays of Truth.

1.2

New Possibilities for Human Motion Studies by Real-Time Light Spot Position Measurement

H. J. WOLTRING

Laboratory of Psychology, University of Nijmegen, Nijmegen

Key Words Motion study Biomechanics · Displacement sensing · Optoelectronics · Transducers Measuring equipment · Light telemetry

Abstract. For the purpose of real-time investigation of human motion in which feedback during a movement must be possible, various optoelectronic techniques for non-contacting motion studies have been compared. Light sources and detectors for single- and dual-axis position measurement are discussed, with special emphasis on multiple light spot sensing. An experimental system based upon one of these detectors is described, and some preliminary results on timing and coordination of human motion are presented.

Introduction

Although the detailed study of human motion has traditionally taken place in medically oriented fields such as physiology and surgical rehabilitation, psychologists have been interested in the control of human motor activities for many years [26]. This interest was partly aroused by practical problems in accurate human motor activities such as aiming, typewriting, morse-keying and assembly tasks, since knowledge of the underlying control processes might result in optimized task design and training procedures.

Unfortunately, such research has often been hampered by the use of rather global measures such as delay and movement times or terminal accuracy. A typical example is the choice reaction task [23], where one of

Received: July 18, 1974.

several stimuli is presented and where a corresponding response is required from the subject. Since the latency and movement times vary with the number of possible stimuli and responses and with the complexity of their interrelationships, various decision and control processes have been inferred from temporal data. It would seem that such processes might be more clearly elucidated by detailed studies of movement trajectories, and by investigating derived quantities such as speed, acceleration and forces during a movement.

Our interest is especially directed towards the timing and coordination aspects of movement, such as may be exemplified with the standard paradigm of manual dotting between alternate points. According to some published results [3], such movements seem to consist of a ballistic phase of an intermittent control nature, followed by a comparatively continuously controlled 'homing' phase to the terminal points.

The possibility of real-time feedback would seem to promise more insight into such control processes. For instance, the relationship between the ballistic and homing phases referred to above might be investigated by turning off the goal stimulus (e.g. a light source) at specific movement phases such as a particular percentage of the total distance to be covered or maximum speed. Alternatively, choice reaction processes might be investigated by changing a stimulus during a movement, again at specific phases. For such purposes, a comparison has been made between various techniques for detailed real-time movement measurement.

Previous techniques for non-real-time studies of human motion have often been based upon photographic or cinematographic techniques [1]. NEUKOMM [12] recently described an interesting real-time system in which the directions of rubber bands connecting corporal points with angular transducers are measured, but the forces thus exercised upon a subject may unfortunately modify his behaviour in an uncontrollable fashion. Other approaches have included the use of accelerometers, goniometers and ultrasonic equipment.

For real-time investigation of intricate human motion, subjects should be burdened as little as possible, and high spatial resolution is also required; optical approaches avoiding the delays inherent in film processing would seem to be quite appropriate for this purpose. TV systems interfaced to subsequent data-processing equipment allow such measurements to some extent, and recent advances in optoelectronics have resulted in a variety of devices promising improved performance in terms of bandwidth, spatial resolution and identifiability of multiple light spots.

A number of such devices are discussed in this paper, with special emphasis on multiple light spot sensing since the coordination of limb segments during motion is of some importance. Multiple position studies may also entail increased reliability of position data, especially in three-dimensional work a single light source will move together with the skin around the subcutaneous bone structure, and will be at some distance from the longitudinal axis of this structure, the mean position of two diametrically mounted light sources on a rigid cuff will coincide more accurately with this axis. Various single- and dual-axis sensors will be described which allow the measurement of a light spot position along a line or in a plane, by the combination of a number of such sensors, three-dimensional studies may be possible as well [24].

A system based upon one of the position sensors to be described has recently been constructed in our laboratory, and some preliminary results on dotting between alternate points are presented at the end of this paper; future efforts will also be directed towards three-dimensional motion studies.

Measurement Techniques

In this section, the transmitting and receiving ends of some optical position measuring systems and their constituents will be discussed A classification has been adopted for active and passive light sources, and for addressable versus non addressable position detectors

Position Marking

Passive versus active light sources The positions to be sensed may be marked by either active or passive light sources Stroboscopy is a well known approach when passive light spots are used the spatial positions of small reflectors attached to the human body must then be measured Passive light sources have an advantage over active sources in not requiring connections to a power supply near the subject, but they entail the following drawbacks firstly, if visible light is used, requirements of contrast may render the experimental situation unsuited for psychological investigation if the subjects must work in semi-darkness in order to prevent measurement disturbances by the ambient light Such artificial circumstances may be acceptable in other fields of investigation, but they will be likely to change the subjects' 'normal' behaviour, especially with intricate movements Secondly, if more than one light spot is to be measured, identification problems arise In three-dimensional or real-time studies especially, this may become quite cumbersome, and some sort of position coding or labelling should be preferred This might be achieved by taking different colours for the various passive light sources if colour television is used at the receiving end, but later in this article it will be explained why standard TV systems are not very well suited for fast and multiple-position studies in a real-time en-

vironment, whilst non standard TV systems become technically rather complex and expensive Pattern recognition techniques (see below) are an alternative approach, but active light sources may in addition be identified by multiplexing techniques in the time or frequency domains

Active light sources LEDs A more promising approach is based upon currently available power LEDs or light emitting diodes having the advantages of (1) being good point sources and having small dimensions, (2) radiating exclusively in the near infrared, and (3) having fast switching characteristics Ordinary miniature incandescent light bulbs fail in these respects, and current developments suggest that optical power for infrared LEDs will increase considerably in the coming years (at present some 100 mW when pulsed at 10% duty cycle)

Since LEDs may be good point sources, the resolution of measuring equipment based upon them will generally be limited by the receiving end The infrared feature of such systems implies that a subject is not distracted by light emanating from his body, e g when coordination studies are carried out under various levels of illumination If the receiving end is sensitive to the ambient light, infrared filters may be applied in front of the detectors, if this does not suffice, one may resort to standard fluorescent light rather than incandescent light for ambient illumination the former type of light does not radiate in the infrared

The fast switching times of LEDs – below 1 μsec – allow both the possibility of labelling by time-multiplexing and of increased signal to noise ratio (SNR) for white receiver noise – as may for instance be expected for silicon photodiodes – the SNR varies inversely with the pulsing duty cycle if the optical energy per cycle is kept constant

Telemetric control and fibre optics A disadvantage of active light sources affixed to the subject is the need for connections to some power supply and control circuitry, as already indicated above This problem may be solved to some extent by the use of telemetric control systems such as described by NEUKOMM *et al* [13], an alternative approach, also suitable in cases where even small light sources such as LEDs are too bulky – e g in studies of finger or animal motion – might be to introduce optical fibres [11] between a central light source and the various points to be monitored Time-multiplexing might be performed either by using distinct light sources per marked position as above, or by means of a small mechanical chopper passing the light towards one fibre exit aperture after another

Position Sensing

In the closely related field of optical tracking [19], a distinction is commonly made between open-loop and closed-loop trackers In closed-loop tracking, the output of the optoelectronic detector is used as an error signal for mechanical realignment of the tracking system onto a single target Such an approach is impractical for high speed, multiple light spot monitoring since one camera per light spot would be required, we will limit ourselves, therefore, to mechanically fixed systems permitting single and multiple light spot sensing

Position sensors may be categorized as (1) 'addressable sensors' in which image points or areas must be addressed and read, like storage locations in a computer memory, and as (2) sensors directly yielding position-dependent output signals caused by the *total* incident light distribution In most cases, some optical system is

required for imaging object space onto the detector surface, and signal-background discrimination may be conducted by optical filtering or by multiplexing techniques.

A typical error source in light-based position sensing is the occurrence of unintentional light reflection; thus, false images may be projected onto a detector surface. Addressable sensors are not very susceptible to such errors, since one may select sectors of object space on the basis of predicted light source positions, whereas non-addressable detectors do not allow such spatial filtering. The use of narrow-angle light sources may prevent reflection errors, but in that case the light source orientation must be controlled; for human motion studies this may be a disadvantage.

Addressable Sensors

These sensors may be subclassified as sequential and random access sensors; in the former, the image plane is scanned in some constant and systematic fashion, which may be a characteristic of the detector proper or a consequence of the external addressing circuitry.

Sequential access systems. A TV camera tube in itself is a random-access sensor; when used in conjunction with standard TV equipment, one obtains a sequential access system [2, 7, 16]. The camera tube deflection signals may be digitized and transferred to subsequent equipment as soon as the scanning electron ray finds a light patch; conversely, the deflection may be crystal controlled for increased accuracy and stability. Such TV oriented systems have the disadvantages of being rather slow and of having low resolution compared to other devices to be described below: some 16–20 msec is required for scanning the whole image plane if interlacing is discarded, at the cost of even more loss in resolution. Since this measuring time would appear to be about the maximum sampling interval allowed for human motion studies – many people consider higher sampling frequencies better suited – identification by time-multiplexing cannot be implemented and all light sources must be turned on simultaneously. Identification must here be done with colour-labelling (CTV) or with computer-controlled prediction or pattern-recognition techniques [8]. Moreover, the resolution of standard TV systems is rather low if one is interested in derived quantities such as speed and acceleration. Because of the slowness of standard TV systems, some form of light chopping is required in order to prevent light-smearing over the image area. Thus, either LEDs may be used – necessitating rather sensitive TV camera tubes at present – or some form of mechanical light chopping at the receiver end must be implemented. Since TV cameras commonly work in the charge-storage mode, light information is preserved until read by a scanning electron ray.

The slowness and low resolution of standard TV systems do not apply to the single-axis equivalent: sequentially addressable, silicon photodiode arrays can already attain resolutions of 1:1,024 at 1 in length [17] and may be read within 50 μsec (minimally), depending on the incident light intensity. Two- and three-dimensional studies require at least two and three single-axis sensors, respectively, each equipped with cylindrical optics. The speed and sensitivity of such sensors would seem to be compatible with time-multiplexed LED properties, but the costs are quite high and cylindrical optics of sufficient quality are hard to obtain.

The strategy of line-by-line scanning is the cause of the slowness of standard TV systems. Although present efforts in solid-state imaging techniques are focussed

mainly on sequential access sensors for TV applications, one might envisage dual-axis photodiode arrays which may be read either column-wise with all rows in parallel, or row-wise with all columns in parallel. Light sources in the time-multiplexed mode might thus be measured in sub-millisecond intervals by scanning all rows simultaneously and subsequently all columns; the required optical system would be standard spherical.

Random access systems. These are often based upon image dissector tubes [5–7, 14, 27] which allow a higher resolution than ordinary TV camera tubes such as Plumbicon or Vidicon. Sensitivity can be very high if photomultipliers are used as detectors behind the dissecting aperture. When employed for single-point monitoring, such systems are effectively electronic closed-loop trackers: the light intensity of a small area of the camera image plane is used as an error signal for deflection control. Both deflections are alternately readjusted in the time-multiplexed mode in the case of dual-axis tracking. For multiple-point monitoring systems [7], previous light source locations must be memorized in order to allow fast search to the current positions.

A problem with some commercially available instruments is that targets must have a special pattern and orientation. If they fail in these respects, resolution will be reduced or the system will not function at all.

Defocussing due, for instance, to distance variations is a common problem for all addressable sensors. For random-access systems, one may try to find the image maximum; for sequential-access systems, some average value must be assessed if more than one 'address' contains light-induced information.

One may conclude that standard TV systems are not very well suited for real-time, multiple-position sensing, unlike single-axis or random-access systems. These, however, may still cause problems in terms of specialized optics or search difficulties. Some of the sensors described in the next section are more promising in these respects.

Non-Addressable Sensors

This class of position-sensitive detectors has the property that position information generated by the total incident light distribution can be derived from a limited number of output signals; information may be available in analog or digital form.

Duo- and quadrant photodiodes. These detectors consist of adjacent diodes on a common substrate (fig. 1); they are often used as position sensors in closed-loop trackers. A properly defocussed light source image is projected onto the various photocells, and the output currents vary with light patch shape, position and intensity. Intensity dependency may be cancelled, for example by taking the ratio of the difference and sum of the output currents from opposite diodes. Although very fast and sensitive, such detectors are unfortunately less promising for open-loop sensing since it is difficult to maintain constant light patch shape and dimensions.

Lateral photodiodes. Photodiodes based upon lateral or longitudinal photo-effects [9, 15, 22, 25] are more suitable: if a p^+-n junction with floating p^+-side is non-uniformly illuminated, a lateral potential may be observed along the junction at the n-side since the common transverse photopotential varies with the incident light distribution [22]. The functioning of p-(i-)n or metal-semiconductor photodiodes having one or two resistive electrodes with properly placed lateral, ohmic contacts is

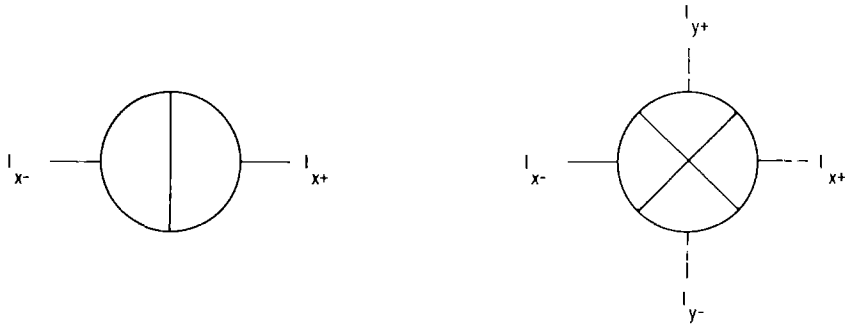


Fig. 1. Duo- and quadrant diodes for closed-loop tracking.

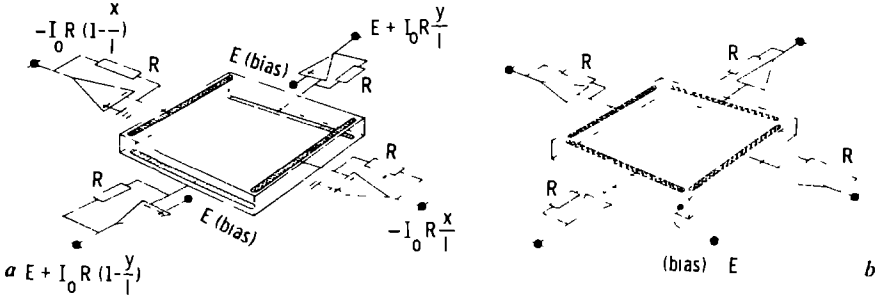


Fig. 2. a Linear lateral detector. *b* Non-linear lateral detector.

closely related to this effect (fig. 2). Basically, the output currents are determined by the junction properties and by the resistive path between the lateral contacts and the light spot position. For the fully reverse-biased case with zero loading impedances shown in figure 2a, the signal currents are linear functions of both light patch power and position along a single axis. The integrated pulse response of such sensors has been derived [9], and it is easy to prove that this quantity corresponds to the steady-state response. By taking the ratios of current differences and sums per axis, energy dependency is again cancelled and two functions are obtained, each linear in only one coordinate.

Such continuous position-sensitive detectors – with non-linearities $<1\%$ – have been used in nuclear physics for particle incidence position and energy measurement [9, 15], but corresponding optoelectronic detectors are not yet available. A slightly different configuration does exist, however [4, 21]. In this case, four lateral contacts with small separating gaps at the vertices are placed on a resistive electrode of a square junction, whilst the other, highly conductive electrode is either grounded or connected to a biasing voltage (fig. 2b). Resolution of such tetra-lateral detectors may be extremely good: up to $1{:}10^4$/axis, depending on the incident light power. Their position characteristics have been analysed [25]; the log ratio of output

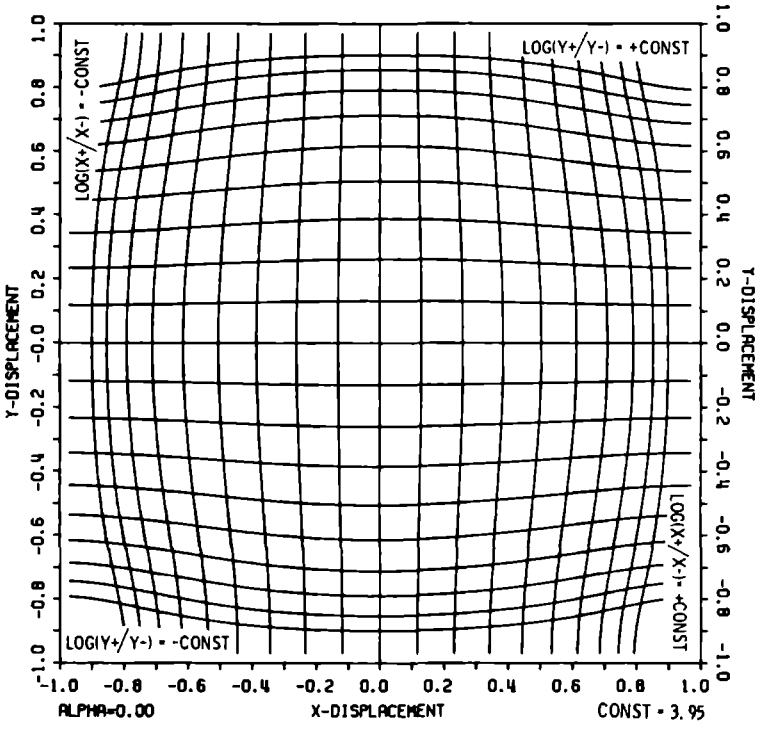


Fig. 3. Fully reverse-biased, tetra-lateral photodiode with zero loading impedances. Loci for constant values of the log output currents ratio per axis LOG(X+/X–) or LOG(Y+/Y–), for equidistant fractions of the log ratio CONST corresponding to the 90% off-centre axis point.

currents per axis has been found to be quite linearly related to the light spot position along the corresponding axis (fig. 3) for the case of zero loading impedances in both the unbiased and fully reverse-biased modes.

Under fully reverse-biased conditions, the output currents of both detector types are linear functions of the incident light distribution; this entails the possibility of discriminating weak light pulses from a stronger but steady background illumination by measuring the output current variations. Since the output currents correspond to the 'average' light patch position, lateral detectors are less susceptible to defocusing than addressable sensors.

A time-multiplexed system based upon the tetra-lateral detector has been described [10] and will soon be commercially available [18]; a simplified version has been constructed in our laboratory.

Spherical lateral detector. A very interesting detector displaying a lateral photoeffect has been reported [20]; at the cost of reduced sensitivity, such detectors

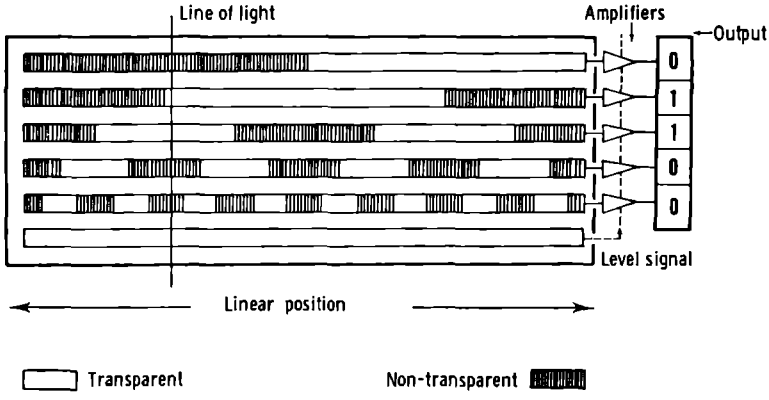


Fig. 4. Binary Gray-coded detector.

would not require focussing optics. A spherically shaped semiconductor crystal is equipped with a photosensitive junction all over its surface. Because of shadowing effects, the junction is non-uniformly illuminated when irradiated by a single, small light source, and lateral photopotentials for single- and dual-axis sensing may be observed through suitably chosen surface contacts. Future developments might render such detectors practical.

Binary encoded detectors. For automatic data processing, e.g. on-line in a small computer, some form of digitizing position data would generally be required. AD-conversion is appropriate for continuous or pulsed signals like in the above cases, but a single-axis sensor directly yielding encoded data may be envisaged (fig. 4). A set of slender photocells next to each other is covered with a binary pattern, and the position of a well-focussed line of light perpendicular to these cells is available in coded form.

Errors due to defocussing may be rendered less serious if a Gray or reflected binary code is used rather than standard binary code; conversion to the latter code is quite easy. A nine-bit Gray-coded detector has been manufactured [4]; an advantage over continuous position sensitive detectors is the reduced effect of noise and false images, but sensitivity to defocussing and the necessity of cylindrical optics are disadvantages as in the case of single-axis, addressable sensors.

It is suggested in conclusion that lateral photodetectors are well suited for real-time and multiple-position sensing if reflection errors are negligible; in other circumstances, digitally coded detectors, random-access systems or single-axis, sequential-access systems would seem to be appropriate.

Experimental Results

An experimental system containing a tetra-lateral photodiode [21] has been constructed and is shown in figure 5. The detector output signals are

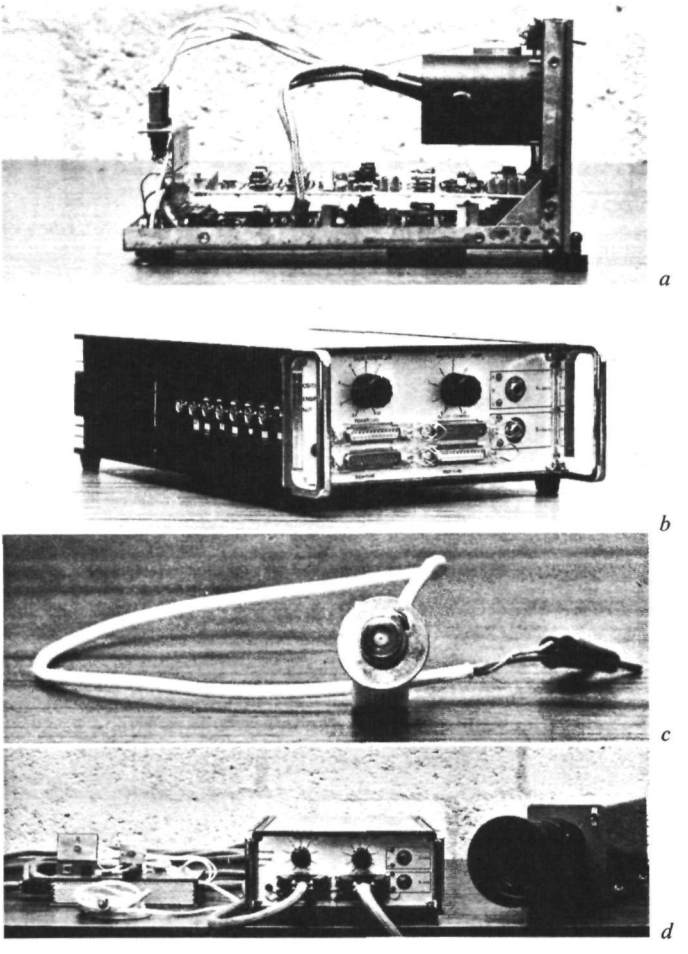


Fig. 5. Experimental position measuring system. *a* Camera unit. *b* Main control unit. *c* LED light source. *d* Overall view.

amplified and fed into a minicomputer either directly or after hardware log ratio conversion with analog circuitry. The system allows time-multiplexing of 3 LEDs; the maximum sampling frequency of 300 Hz for a 10% LED duty cycle is limited by the processing electronics, since the settling time of the detector output currents may be in the order of 30 μsec.

Only one detector is currently under investigation; its position charac-

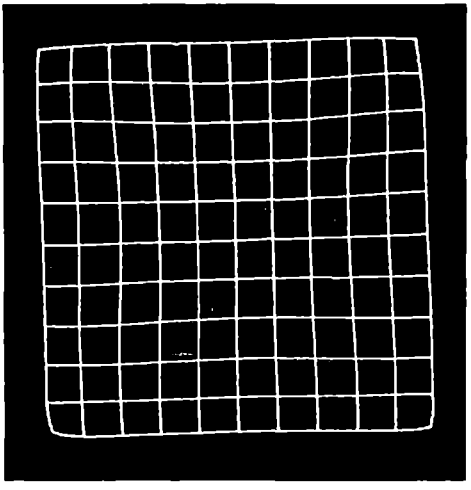


Fig 6 Experimental position characteristics of an unbiased tetra-lateral photodetector with zero loading impedances (log ratio of output currents per axis for equidistant, rectilinear displacements of a light spot image over the detector surface).

teristics are shown in figure 6. The position-sensitive area is partly masked in order to prevent a light spot being too closely incident to a lateral contact. In this fashion, all four output currents keep reasonable values with respect to the detector noise currents if a light spot image of constant power is moved across the unmasked area. The experimental curves do not cover the whole unmasked area, since the particular optical system used allowed an image field diameter approximately equal to the side length of the – square – mask.

The curves in figure 6 correspond to the inverse of figure 3; in the unbiased mode, the latter become slightly more rectilinear and more equidistant, at the cost of some signal loss because of internal leakage [25] and increased settling time. The shapes of both figures suggest that linearity may be improved by curving the lateral contacts slightly, and optical correction might serve the same purpose.

The deviations at the edges of the experimental curves may be due to defocussing or to a larger spacing between adjacent contacts than is admissable for the theoretical model [25], but the various non-symmetrical deviations suggest that non-uniformity of the resistance layer is involved. In the present detector, this layer is formed by the non-depleted part of

the bulk silicon in a Schottky-barrier photodiode, and the uniformity of such layers is known to be poor. Better conformance with theory may be expected for diffused or ion-implanted resistance layers in semi-conductor material of higher resistivity, as has already been reported for nuclear position detectors.

Resolution depends upon the incident light power: a value of better than 1:400 has been observed along a line in the field of view when using wide-angle LEDs with an estimated peak optical power of 25 mW at 1 m distance from the camera. The optical system had an aperture of 5 cm diameter. Resolution is expected to increase by well over a factor 4 for distances up to 3 m by using faster processing electronics and reverse biasing: background illumination effects are cancelled in the present set-up by optical filtering, and detector speed may increase by a factor 2 or more because of decreased junction capacitance. The comparatively long measurement delay of 300 μsec does not allow higher optical power at the time of measurement since internal heating of the LEDs causes optical power decay during a light pulse. For shorter pulses the thermal capacity of the LEDs may be sufficiently large to keep power decay between reasonable limits.

Applications

The results of a simple dotting task are shown in figure 7. A subject was alternating with a pencil between two targets of 5 mm width at 10 cm seperation, with an LED mounted close to the point of the pencil. No real-time requirements were made for this simple experiment, and data smoothing and differentiation were performed by discrete Fourier transformation techniques after acquisition of a complete record with 256 samples/axis. The adverse effects of finite record length were reduced by searching for two equal data points per axis from the record ends; the outer record parts were set equal to these values prior to smoothing and differentiation. The smoothing function was a cosine curve up to a cut-off frequency corresponding to the first quarter period.

Position and velocity data have thus been obtained for sampling frequencies up to 80 Hz, but the increased effects of measurement noise on differentiated signals require smoothing until approximately 10 Hz for acceleration data. Signal deformation due to the non-linearity of the position detector is of a global rather than a local nature, since the lines in figure 6 display only a gradual curvature.

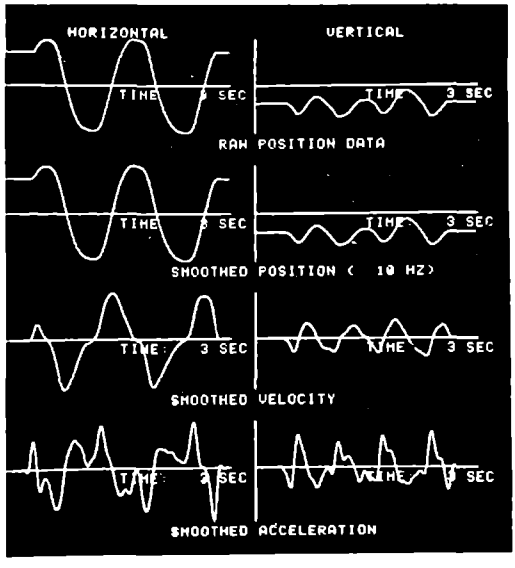


Fig. 7. Position, velocity and acceleration in a dotting task (smoothing cut-off frequency: 10 Hz).

The curves in figure 7 show different shapes in the left- and right-ward motions: this was probably caused by masking of the left target by the subjects' moving left hand. The horizontally extended and vertical motion components seem to be of a similar nature, with only temporal differences: the horizontal component virtually terminates above the target and is followed by a continued vertical movement until target attainment. It would appear that the subject terminates this vertical component by simply coming in contact with the target.

An intermittent control strategy operating at a frequency of approximately 10 Hz and a distinct homing phase of a comparatively continuous nature have been suggested in the literature [23]; their existence cannot be derived from the present data and more definite results are expected from an improved measuring system as outlined above.

Conclusions

Contemporary optoelectronics allow real-time investigation of human motion with a minimum of hindrance to the subject, by automatic posi-

tion sensing in one-, two- or three-dimensional space of small reflectors or active light sources attached to the human body. The light signal may be separable from background illumination by means of optical filtering and by multiplexing techniques; the latter allow an easy implementation of multiple light spot sensing when active light sources are used. Subjects cannot be distracted from their task by the emanating light if infrared light sources are used as position markers.

These approaches avoid the elaborate data reduction processes of classical film analysis and, in particular, give promise of the possibility of real-time feedback as an experimental tool in motion studies.

The position detectors discussed in this article allow single- and dual-axis position sensing; by suitable combinations of independently positioned detectors, three-dimensional motion may also be investigated.

Acknowledgements

The author wishes to thank Mr. H. J. J. JANSSEN for the construction of a position measuring system based on a tetra-lateral photodetector, Mrs. D. DIJCKHOFF-LAENEN for the drawings in this paper, Dr. D. COTTON for linguistical suggestions, and Dr. E. G. J. EIJKMAN for valuable comments on an earlier draft of this article.

References

1 BERNSTEIN, N.: The coordination and regulation of movements (Pergamon Press, Oxford 1967).
2 Colorado Video Inc.: Trade literature (Boulder, Colo.).
3 CROSSMAN, E. R. F. W. and GOODEVE, P. J.: Feedback control of hand movement and Fitts' Law. Commun. Experimental Psychology Society (1963); see WELFORD [23].
4 Electro-Nuclear Laboratories: Trade literature on Gray-coded and photovoltaic linear detectors (Infrared Industries, Waltham, Mass.).
5 EMR-Schlumberger: Trade literature on the Optical Data Digitizer (Vélizy Villacoublay, France).
6 FITTON, G. M.: The Optical Data Digitizer – an eye to the future. Seminar on quantitative imagery in the bio-medical sciences II. Soc. Photo-Opt. Instr. Eng. 17th Annu. Techn. Meet. and Equipment Display, San Diego, Calif. 1973.
7 Hamamatsu TV Co. Ltd: Trade literature on physical motion analysis and measuring TV-systems (Ichinocho, Hamamatsu, Japan).
8 INGEN SCHENAU, G. J. VAN: Arm motion study by means of television interfaced to a digital computer. (In Dutch.); MSc. thesis, Department of Applied Physics, Technical University of Delft (1973).

9 KALBITZER, S. and STUMPFL, W: A nomogram for the design of position sensitive silicon detectors Nucl. Instr. Meth *77* 300–302 (1970).
10 LINDHOLM, L.-E. and OEBERG, K E. T.: An opto-electronic instrument for remote on-line movement monitoring Biotelemetry *1* 94–95 (1974)
11 MILLER, S. E; MARCATILI, E. A., and LI, T.: Research toward optical fiber communication systems Proc. IEEE *61* 1703–1751 (1973).
12 NEUKOMM, P. A: The rubber band goniometry. A telemetric method for the measurement of angle, angular velocity, displacement and velocity Biotelemetry *1* 12–20 (1974).
13 NEUKOMM, P A; GRANDJEAN, R.; WALSER, K, and WALTER, H.: A radio-controlled light-pulse equipment for the photographic measurement of human body movement Biotelemetry *1* 94 (1974)
14 Optron, Division Universal Technology, Inc: Trade literature on non-contacting electro-optical measurement systems (Woodbridge, Conn.).
15 OWEN, R. B. and AWCOCK, M L One and two dimensional position sensing semiconductor detectors. IEEE Trans Nuclear Sci NS-*15* 290–303 (1968).
16 Prosthetics Control Laboratory, Department of Applied Physics, Technical University of Delft, the Netherlands.
17 Reticon Corp · Trade literature on solid state line scanners (Mountain View, Calif.).
18 Selcom, AB: Trade literature on the Selspot system (Molndal, Sweden).
19 Special issue on optical tracking. Appl. Opt. *5* 481–532 (1966).
20 TORE PERSSON, B and STROMQVIST, H. O.· A spherical semiconductor photocell using lateral photoeffect. IEEE Trans. Electron Devices ED-*15* 129–134 (1968).
21 United Detector Technology, Inc.. Trade literature on continuous position sensitive or Posicon photocells (Santa Monica, Calif).
22 WALLMARK, J. T.: A new semiconductor photocell using lateral photo-effect. Proc. IRE *45* 474–483 (1957).
23 WELFORD, A. T.: Fundamentals of skill (Methuen, London 1968).
24 WOLTRING, H. J · 3-Dimensional light spot position and movement measurement – computational aspects. Report 74AO02 (Laboratory of Psychology, University of Nijmegen, Nijmegen 1974)
25 WOLTRING, H. J.: Single- and dual-axis, lateral photodetectors of rectangular shape (to be published).
26 WOODWORTH, R S: The accuracy of voluntary movement Psychol. Rev. Monogr. Nr. 13 (1899).
27 Zimmer OHG: Trade literature on the Electro Optical Displacement Transducer (Rossdorf, FRG).

Request reprints from. H. J. WOLTRING, Laboratory of Psychology, University of Nijmegen, Erasmuslaan 16, *NL-6804 Nijmegen* (The Netherlands)

IEEE Transactions on Electron Devices
*22:*581-590 (1975)

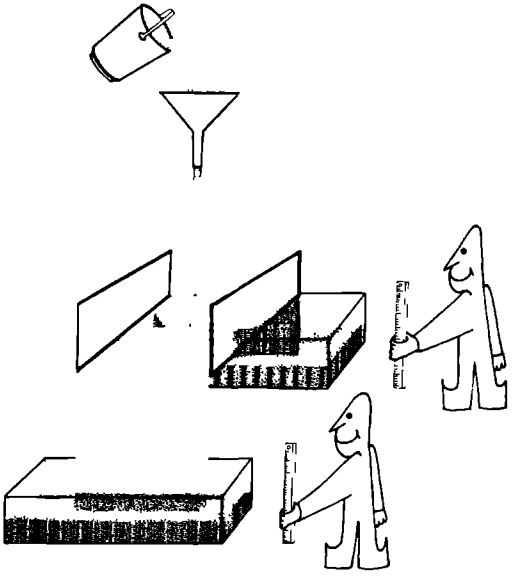
Le goût de l'exactitude, l'impossibilité de se contenter de notions vagues, de s'attacher à des hypothèses, quelques séduisantes qu'elles soient, le besoin d'apercevoir clairement la liaison des propositions et le but où elles tendent, sont les fruits les plus précieux de l'étude des mathematiques.

S.F. LaCroix

1.3

Single- and Dual-Axis Lateral Photodetectors of Rectangular Shape*

HERMAN J. WOLTRING



Hydraulic model of the single-axis duo-lateral photodiode with zero terminating impedances in the fully reverse-biased mode

* A number of position characteristics which were not contained in the published paper have been added as appendices 1, 2 and 3

***Abstract*—A 2-dimensional diffusion model of rectangular single- and dual-axis, lateral photodiodes is analyzed and the different underlying lateral effects are discussed for the small-signal (unbiased) and fully reverse-biased modes, with either infinite terminating impedances at midlateral point contacts or zero loading impedances at extended lateral contacts. For the single-axis detector with two opposite extended contacts, previous results of the 1-dimensional model appear to remain valid, and the output current difference to sum ratio in the fully reverse-biased mode is a linear function of the position of a narrow incident light beam. For the dual-axis detector with extended contacts at all four sides—except for small gaps at the vertices—an approximately linear relation has been found between the light spot position along an axis and the log ratio of the corresponding output currents. The transient responses of these configurations are discussed, and the utility of operation in the small-signal mode is illustrated with some typical parameter values.**

I. INTRODUCTION

UNDER nonuniform illumination of a p-n junction, a photopotential may be observed along the junction barrier since the common transverse photopotential varies with the incident light distribution This *lateral* or *longitudinal* photo-effect—the latter term corresponds to Russian terminology—may be exploited for position measurement of incident radiation; it was first reported in 1930 by Schottky [1], who used a rectangular Cu–CuO_2 metal–semiconductor junction of a few cm^2 area with a single extended contact on the semiconductor n-layer in parallel with one of the detector sides. The contact was short-

Manuscript received December 5, 1974; revised March 5, 1975.
The author is with the Laboratory of Psychology, University of Nijmegen, Erasmuslaan 16, NL-6804 NIJMEGEN, The Netherlands.

circuited to the metal face, which behaved as an equipotential layer due to its low resistivity A line of light was projected onto the detector surface in parallel with the contact, and the photocurrent through the external circuit was found to decrease with increasing distance between the contact and the line of light Depicting the detector as a resistive and leaky telegraph line, Schottky presented the correct theoretical interpretation for this phenomenon the transverse photopotential difference and the current towards the lateral contact are governed by the diffusion process in both junction faces and by the charge carrier transport through the rectifying junction For small transverse potential differences, the latter process may be linearized, and an analytical solution was presented

Schottky's work was reported for its theoretical implications, it passed into oblivion until Emmons [2] brought it to attention again in 1967 Eleven years before, Wallmark [3] had rediscovered the lateral effect and extensive research has been reported since Unlike Schottky, Wallmark used a Ge–In p^+-n junction with a floating p^+-layer and point contacts to the n-layer, he measured the potential difference between two of these contacts which were separated by approximately 1 mm Under such conditions, external currents are not involved, and Wallmark explained the lateral potential difference in terms of a 1-dimensional model for the internal charge carrier transport, which had come to be called the *surface recombination* or *barrier reinjection* effect [4] photon-induced separated carriers diffuse along both junction faces and recombine in the p-n barrier at distance remote from their origin of separation Lucovsky [5] presented a generalized analysis in which the photopotential across the junction was expressed in terms of two partial differential equations which covered the diffusion and recombination processes for the steady-state and small-signal time-dependent cases Subsequent authors have extended these analyses to cases where the influences of boundary shape, external impedances, and biasing potential modify [6], [7] or virtually replace [8], [9] Wallmark's explanation of the lateral photo-effect Thus Allen [6] analyzed the behavior of a circular plane diode with four lateral spot contacts at the ends of two perpendicular axes, Taubkin and Frimer [7] investigated a zero-bias 1-dimensional model with a finite resistance between two sensing contacts, and LaBaw [8] presented a model for a dual-axis square photodiode under fully reverse-biased conditions with zero impedances at extended lateral contacts along the four boundary sides

These extensions of the Lucovsky theory were all steady-state models, unlike Connors' [9] presentation of the 1-dimensional model with finite terminating impedances for the small-signal and fully reverse-biased modes In his analysis of the time-dependent and steady-state cases, he emphasized the different effects involved in a) Wallmark-type photodiodes with floating junction and negligibly large terminating impedances, and b) fully reverse-biased photodiodes with finite or even zero loading impedances.

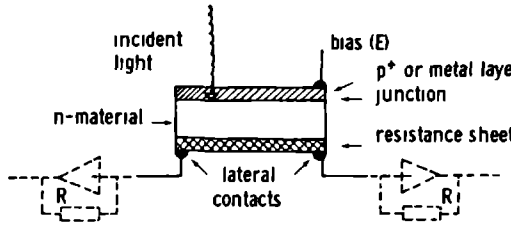


Fig 1. Cross section of a lateral photodiode.

In the latter mode, surface recombination is not involved, and the generated photocurrent is fully available at the joint lateral contacts

Lateral effects in fully reverse-biased junctions with zero loading impedances have been thoroughly investigated in position-sensitive semiconductor detectors for nuclear radiation [10]–[12] Such detectors differ from their optical counterparts only in being optimized for (e g) α, β, or γ radiation by having a graded rather than a surface or Schottky-barrier junction as may be achieved by diffusion or ion-implantation techniques

In this paper, Connors' analysis is extended to the 2-dimensional rectangular junction operated in the small-signal (unbiased) or fully reverse-biased (saturated) modes, this operational limitation allows a more complete theoretical treatment than would be possible for the general nonlinear mode and is motivated by practical considerations with respect to signal–background interaction and position linearity The results that have been published for nuclear position sensitive detectors follow as special cases The theory assumes surface recombination and external current collection as the sole causes of the lateral effect, other phenomena such as tunneling in Schottky-barrier junctions are not considered The charge carrier generation process is not considered in detail, since this is not specific for lateral effect detectors charge carriers may be generated by monochromatic or white light, or by other types of radiation The transit time of the generated charge carriers into the barrier field will be assumed to be negligibly small with respect to the time-constants involved in the distributed RC character of the detector [5], [13], the rectangular area will be assumed to be square with side length l since oblong dual-axis detectors would not seem to have much practical value, and since the length to breadth ratio has no effect in the special case of the single-axis duolateral detector discussed below Generalization to the oblong case is a straightforward operation as may become apparent from the discussion in Section III

For the small-signal and fully reverse-biased modes, the junction may be considered to have a resistive electrode of thickness w_d and resistivity ρ_d (see Fig 1) The depletion layer width may be considered to be constant, which corresponds with a fixed capacitance per unit area C_d, it should be realized that the values of these quantities at

a reverse bias may differ from those at zero bias The resistance sheet may be either the bulk semiconductor material outside the depletion layer or a diffused or ion-implanted resistance sheet in a semiconductor material of much higher resistivity The latter construction may entail increased uniformity of the resistance layer, and thus a better agreement between theoretical and experimental results [10], [12] A second advantage of this approach lies in the possibility of increased barrier width yielding a higher junction sensitivity without deterioration of the aspect ratio l/w_d for a proper lateral effect, this quantity should be much larger than unity [9]

The material at the other face of the junction will be assumed to have a much higher conductivity, i e , we will consider p⁺-n or metal–semiconductor type junctions for which Lucovsky's equations given in (1) and (4) obtain [5] This restriction is unnecessary in the fully reverse-biased mode for the steady-state case, since the lateral effect is then solely determined by the resistance sheet and terminating impedances Current leakage effects at the boundaries will be neglected, since contemporary p-i-n or Schottky-barrier photodiodes generally exhibit low boundary leakage, especially if they are equipped with a guard ring structure Some practical parameter values are presented in Section V

II GENERAL EQUATIONS

Lucovsky's equation [5] for the steady-state lateral effect in a p⁺-n junction

$$\nabla^2 U - \frac{\rho_d J_d}{w_d}\left[\exp\left(qU/kT\right) - 1\right] = -\frac{\rho_d}{w_d} J_s \qquad (1)$$

may be linearized[1] for small potential differences $|U| \ll kT/q$, as in

$$\nabla^2 U - \alpha^2 U = -f_s \qquad (2)$$

where

$U = U(x,y)$	transverse potential difference
J_d	reverse saturation current density
$q/kT = 40\ \mathrm{V}^{-1}$	at 300°K
$J_s = J_s(x,y)$	generated photocurrent density
$\alpha = (\rho_d J_d q/w_d kT)^{1/2}$	Lucovsky's "lateral fall-off parameter" [5], α^2 may physically be interpreted as the product of the lateral sheet resistance per square and the transverse conductance per unit area at zero bias
$f_s = f_s(x,y) = (\rho_d/w_d)\, J_s(x,y)$	normalized source term

In the fully reverse-biased mode one may neglect the exponential term in (1)

$$\nabla^2 U = -(f_s + f_d) \qquad (3)$$

where

$$f_d = \frac{\rho_d}{w_d} J_d$$

Thus steady background illumination and diode leakage current are equivalent in their effects upon the fully reverse-biased lateral diode, whereas J_d in (2) occurs in the parameter α only

For the types of boundary conditions to be discussed in this paper, the potential U in (2) and (3) is a linear function of the forcing source terms f_s or $(f_s + f_d)$ U in (3) may, therefore be split into a steady value U_d caused by external bias leakage current and the constant *background* component of f_s, and into a signal potential difference U_s caused by the *signal* component of f_s [9] In such a fashion, (3) for U_s may be viewed as the case $\alpha = 0$ of (2), which will be done in the remainder of this paper

For the time-dependent case, Lucovsky expanded (3) with a capacitive term

$$\nabla^2 U - \alpha^2 U - \beta \frac{\partial U}{\partial t} = -f_s \qquad (4)$$

where

$\beta = C_d \rho_d / w_d$	elementary time constant per unit area
$U = U(x,y,t)$	time- and place-dependent
$f_s = f_s(x,y,t)$	

The validity of (2) and (4) may be extended to the small signal case for an arbitrary bias Allen [6] and Connors [9] have shown that α^2 must be expanded with a factor $\exp(qU_d/kT)$, since also C_d is a function of U_d, the present analysis which requires α and β to be constants will obtain if U_d may be assumed to be constant over the whole detector area This implies low leakage currents and background illumination levels which are not required for operation in the fully reverse-biased mode (see Connors [9])

For proper exploitation of lateral effects, various contact configurations may be considered (Figs 2–4) Analytical solutions for such configurations would appear to be possible only for infinite terminating impedances at lateral point contacts (i e the local value of U is observed without loading effects) and for zero terminating impedances at extended lateral contacts along the junction sides In such circumstances we find homogeneous boundary conditions of the Neumann type (field strength component normal to the boundary $\partial U/\partial n = 0$) or Dirichlet type ($U = 0$) for (2) and (4)

Although Figs 2–4 show physically rectangular junctions, practical detectors may be given an active area of rectangular shape by the terminating contacts and by a controlled diffusion or ion-implantation process for the creation of the resistance layers Thus Neumann-type boundaries may be approximated at transitions from ion-

[1] It is assumed that the exponential term in (1) allows a good description of the detector's behavior At zero bias, a better approximation is often $\exp(qU/\gamma kT)$ with $1 \leq \gamma \leq 4$, which would correspond with a change in α see Marfaing [14] for a discussion pertaining to lateral effects

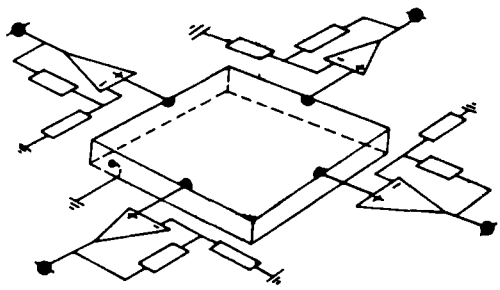


Fig 2 Dual-axis square Wallmark diode with floating junction (bottom view)

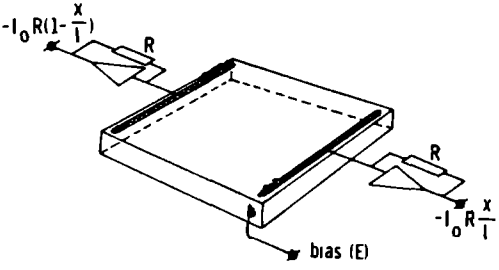


Fig 3 Single-axis duolateral detector (bottom view)

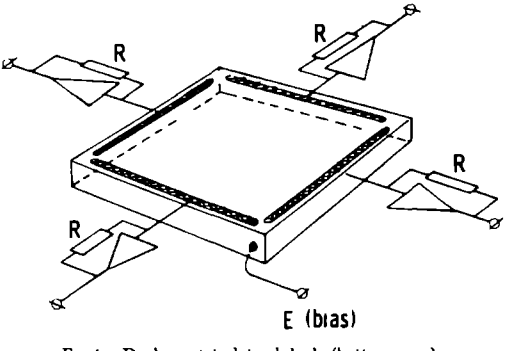


Fig 4 Dual-axis tetralateral diode (bottom view)

implanted to purely bulk material regions, while Dirichlet-type boundaries may occur at the ohmic and short-circuited contacts possibly combined with locally higher implantation doses in order to improve the effective contact rectilinearity

In the next subsections, the position sensitivity to a small light spot will be compared for three cases allowing analytical solutions 1) Neumann conditions at all sides with point contacts at the centers of each side, corresponding to the rectangular equivalent of Wallmark-type diodes (see Fig 2), 2) Neumann conditions at two opposite sides and Dirichlet conditions at the remaining sides (i e, contacts extending along the latter sides with zero loading impedances—see Fig 3), and 3) Dirichlet conditions at all four sides, with negligibly small contact separating gaps at the vertices (see Fig 4) In spite of the different mechanisms involved, the mathematical descriptions for the fully reverse-biased and small-signal modes will appear to be of the same form In all cases, the initial signal potential difference will be taken as zero, and the step response will be presented

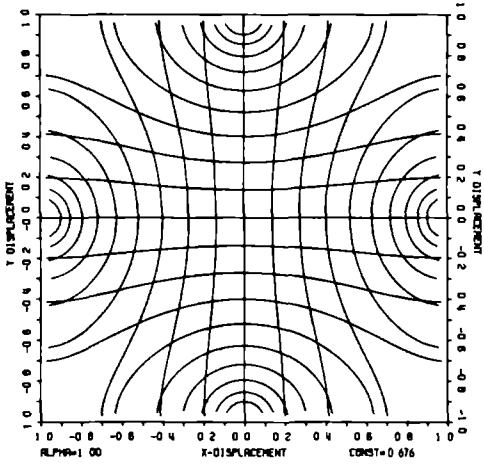


Fig 5 Dual-axis Wallmark diode Loci for constant values of log output potential ratio per axis log $(X+/X-)$ or log $(Y+/Y-)$, for equidistant fractions of log ratio CONSTant corresponding to the 90% off-center axial point

III SOLUTIONS FOR THE RECTANGULAR CASE

Equation (4) is known as the "generalized heat diffusion equation" and corresponds to the transmission-line equation for the case of vanishing inductivity, its solution for various types of boundary conditions may be found in the literature [15], [16] For the given initial and boundary conditions, the response to step signal $f_s(x,y,t) = f(x,y)$, $t > 0$, may be written as a Fourier expansion with terms of the form

$$\frac{\begin{matrix}\sin\\ \cos\end{matrix}\left(m\pi\frac{x}{l}\right)\cdot\begin{matrix}\sin\\ \cos\end{matrix}\left(n\pi\frac{y}{l}\right)}{\alpha^2+(m^2+n^2)(\pi^2/l^2)}\cdot\left[1-\exp\left(-\left[\alpha^2+(m^2+n^2)(\pi^2/l^2)\right]\frac{t}{\beta}\right)\right] \quad (5)$$

The sine should be chosen for Dirichlet-type boundary conditions ($U = 0$) at the ends of the corresponding axis, and the cosine for Neumann-type conditions The weighting factors for this expansion follow in a linear fashion from the generated photocurrent distribution $J_s(x,y,t)$ [16]

For a light spot of small dimensions, $f(x,y)$ may be approximated with Dirac delta functions, at distances which are large with respect to these dimensions, the deviation in U due to this approximation becomes negligibly small,

$$f_s(x,y,t) = F_s \cdot \delta(x - x_0) \cdot \delta(y - y_0), \qquad t > 0 \quad (6)$$

where

$$F_s = I_s \cdot \frac{\rho_d}{w_d}, \qquad I_s = \text{generated photocurrent}$$

For the given initial condition $U(x,y;0) = 0$, the weighting factors for (5) have been found to become

$$\frac{F_s}{l^2} \cdot \frac{\sin}{\cos}\left(m\pi \frac{x_0}{l}\right) \cdot \frac{\sin}{\cos}\left(n\pi \frac{y_0}{l}\right) \quad (7)$$

and the solution for $t \geq 0$ may be expressed as

$$U(x,y,t) = F_s \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \frac{\dfrac{\sin}{\cos}\left(m\pi \dfrac{x}{l}\right) \cdot \dfrac{\sin}{\cos}\left(m\pi \dfrac{x_0}{l}\right) \cdot \dfrac{\sin}{\cos}\left(n\pi \dfrac{y}{l}\right) \cdot \dfrac{\sin}{\cos}\left(n\pi \dfrac{y_0}{l}\right)}{\alpha^2 l^2 + (m^2 + n^2)\pi^2} \cdot \left[1 - \exp\left(-\{\alpha^2 l^2 + (m^2 + n^2)\pi^2\} \frac{t}{\beta l^2}\right)\right] \quad (8)$$

This expression generally yields finite values for U and will be particularized below for the three boundary conditions indicated in the previous section The *normalized* fall-off parameter αl is of some importance, its value in the small-signal mode may, in practice, range between 0 1 and 10 (see Section V, where also some values of the elementary time-constant βl^2 are presented)

A Square Wallmark Diode (Neumann Conditions at all Boundaries Fig 2)

In this case, no external currents occur, and the cosine terms should be chosen in (8) The value of U at the midlateral point contact $(0, \frac{1}{2}l)$ follows as

$$U(0,\tfrac{1}{2}l,t) = F_s \sum_{n=-\infty}^{+\infty} (-1)^n \cos\left(2n\pi \frac{y_0}{l}\right) \cdot \sum_{m=-\infty}^{+\infty} \frac{\cos\left[m\pi(x_0/l)\right]}{\alpha^2 l^2 + (m^2 + 4n^2)\pi^2} \cdot \left[1 - \exp\left(-\{\alpha^2 l^2 + (m^2 + 4n^2)\pi^2\} \frac{t}{\beta l^2}\right)\right] \quad (9)$$

For the steady-state case, the exponential terms cancel, and we find

$$U(0,\tfrac{1}{2}l,\infty) = F_s \sum_{n=-\infty}^{+\infty} (-1)^n \cos\left(2n\pi \frac{y_0}{l}\right) \cdot \sum_{m=-\infty}^{+\infty} \frac{\cos\left[m\pi(x_0/l)\right]}{\alpha^2 l^2 + (m^2 + 4n^2)\pi^2} \quad (10)$$

By virtue of the Fourier expansions for hyperbolic functions [17],

$$\sum_{m=-\infty}^{+\infty} \frac{\cos(mx)}{a^2 + m^2} = \frac{\pi}{a} \cdot \frac{\cosh\{a(\pi - x)\}}{\sinh(a\pi)}$$

and

$$\sum_{m=-\infty}^{+\infty} m \cdot \frac{\sin(mx)}{a^2 + m^2} = \pi \cdot \frac{\sinh\{a(\pi - x)\}}{\sinh(a\pi)} \quad (11)$$

(10) may be written as

$$U(0,\tfrac{1}{2}l,\infty) = F_s \sum_{n=-\infty}^{+\infty} (-1)^n \cos\left(2n\pi \frac{y_0}{l}\right) \cdot \frac{\cosh\{a_n[1 - (x_0/l)]\}}{a_n \cdot \sinh(a_n)} \quad (12)$$

where

$$a_n = (\alpha^2 l^2 + 4n^2\pi^2)^{1/2}$$

The values of U at the remaining midlateral point contacts follow from (12) by considerations of rectangular symmetry x_0 and y_0 may be replaced with $l - x_0$, $l - y_0$, etc

The steady-state position characteristics of this configuration appear to be far from linear they resemble those of the unbounded lateral effect where the steady-state potential difference between contacts is a linear function of the log ratio of contact-to-light spot distances, if these are small with respect to α^{-1} (see Allen [6]) The output signal dependency on the incident light power may be compensated by ratioing output signals when measured with respect to the opposite junction face Among the various choices that may be envisaged for this the steady-state log ratio of output signals per axis has been found to be rather linearly related to the light spot position along that axis for the central diode area (see Fig 5 for $\alpha l = 1$) In this area, the log ratio is less dependent on position variation along the other axis Position sensitivity here varies linearly with $\alpha^2 l^2$ for $\alpha l \ll 1$, and with αl for $\alpha l \gg 1$ Linearity increases somewhat with αl but the

signals actually available at the sensing contacts decrease due to increased surface recombination

The configuration cannot be operated in the fully reverse-biased mode ($\alpha = 0$), since the generated photocurrent would cause a photopotential offsetting the reverse bias until surface recombination occurs Since α varies strongly with temperature (see Section IV), one may conclude that the dependency of the position sensitivity on α renders the square Wallmark-type photodiode rather unsuitable for practical applications, unlike the two configurations to be discussed below in which external currents are involved

B Duo-Lateral Diode (Dirichlet Conditions at Two Opposite Sides, Neumann Conditions at the Remaining Sides Fig 3)

Assuming extended lateral contacts over the whole side at $x = 0$ and $x = l$, (8) becomes

$$U(x,y,t) = F_s \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \frac{\sin\left[m\pi(x_0/l)\right]\cdot\sin\left[m\pi(x/l)\right]\cdot\cos\left[n\pi(y_0/l)\right]\cdot\cos\left[n\pi(y/l)\right]}{\alpha^2 l^2 + (m^2+n^2)\pi^2} \cdot\left[1 - \exp\left(-\left\{\alpha^2 l^2 + (m^2+n^2)\pi^2\right\}\frac{t}{\beta l^2}\right)\right] \quad (13)$$

The current I_{x-} through the contact at $x = 0$ follows as

$$I_{x-} = I \quad (x = 0,t) = \frac{w_d}{\rho_d}\int_0^l \left.\frac{\partial U}{\partial x}\right|_{x=0} dy = I_s \sum_{m=-\infty}^{+\infty} \frac{m\pi\cdot\sin\left[m\pi(x_0/l)\right]}{\alpha^2 l^2 + m^2\pi^2} \cdot\left[1 - \exp\left\{-\left(\alpha^2 l^2 + m^2\pi^2\right)\frac{t}{\beta l^2}\right\}\right] \quad (14)$$

since all terms with $n \neq 0$ cancel Thus the y-coordinate is no longer involved[2] This implies that the transient and steady-state results of Connors' 1-dimensional analysis [9] remain valid for a light spot rather than a long *line* of light in parallel with the y-axis For the steady-state case we find, by using (11),

$$I_{x-} = I_s \cdot \sum_{m=-\infty}^{+\infty} \frac{m\pi\cdot\sin\left[m\pi(x_0/l)\right]}{\alpha^2 l^2 + m^2\pi^2} = I_s \cdot \frac{\sinh\left\{\alpha(l - x_0)\right\}}{\sinh(\alpha l)} \quad (15)$$

[2] This is equally true for the oblong detector for which $U(x,y,t)$ may be written as

$$U(x,y,t) = \frac{F_s}{W}\cdot\sum_{m=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty} \frac{\sin\left(m\pi\frac{x_0}{l}\right)\cdot\sin\left(m\pi\frac{x}{l}\right)\cdot\cos\left(n\pi\frac{y_0}{l'}\right)\cdot\cos\left(n\pi\frac{y}{l'}\right)}{\alpha^2 + \left(\frac{m^2}{l^2} + \frac{n^2}{l'^2}\right)\pi^2} \cdot\left[1 - \exp\left\{-\left[\alpha^2 + \left(\frac{m^2}{l^2} + \frac{n^2}{l'^2}\right)\pi^2\right]\frac{t}{\beta}\right\}\right]$$

with diode dimensions l,l' along the x- and y-axes, respectively In particular, the elementary output current time-constant βl^2 does *not* depend on the contact length l'

Two special cases may be distinguished on the basis of this relationship for $\alpha l \to 0$ (fully reverse-biased mode), (15) may be linearized and becomes

$$\lim_{\alpha l \to 0} I_{x-} = I_s \cdot \left(1 - \frac{x_0}{l}\right) \quad (16)$$

while, for the case $\alpha l \to \infty$, in which surface recombination prevails,

$$\lim_{\alpha l \to \infty} I_{x-} = I_s \cdot \exp(-\alpha x_0) \quad (17)$$

Thus the fully reverse-biased duolateral photodiode with zero loading impedances displays, the interesting property of a linear relation between output current and light spot position along the pertinent axis The same property has been derived for the integrated pulse response of duolateral semiconductor detectors in nuclear physics [10]–[12], which is formally equivalent to the steady-state response

Under fully reverse-biased conditions, the lateral effect is not caused by diode surface recombination but by the current distribution towards the lateral contacts only, and the properties of the junction face opposite to the resistance sheet have no effect upon the steady-state behavior A dual-axis duolateral linear position sensitive detector [10] may, therefore, be visualized with a single-axis detector per junction face (see Fig 6) Power dependency is cancelled by taking the ratio of, e g , current difference and sum per axis Symmetrical defocusing such as may approximately occur when an optical imaging system is used at not too large angles of incidence, will have no influence because of the linearity property of (16), this is of some use in practical applications of lateral photodiodes [18]

Similar 2-dimensional analyses of the single-axis duolateral photodiode have been presented by Buzanova *et al* [19] for the steady-state case, and by Vasil'yev [20] for the transient case Both authors present a small-signal analysis and neglect the influence of surface recombination, corresponding to $\alpha = 0$ Since in the case of Buzanova *et al* [19], αl appears to range between 0 13 and 0 25, this is an acceptable approximation the theoretical error with respect to (15) varies between 0 and 1 percent which is consistent with their experimental results

Practically attainable linearity of duolateral detectors may be quite good Gigante [12] describes an oblong single-axis silicon detector with a surface barrier and lithium drifted resistance layer exhibiting 0 05-percent

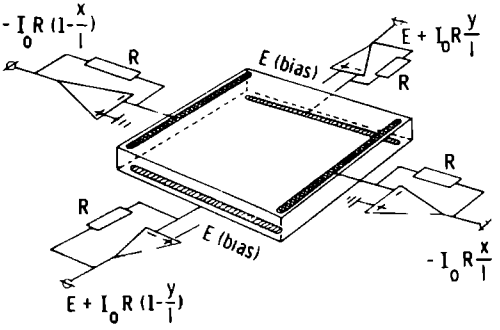


Fig. 6. Dual-axis duolateral diode.

nonlinearity when used for position measurement of incident nuclear radiation.

For large values of αl, surface recombination prevails and (17) implies that the log ratio of output signals is a linear function of position, as was found to be approximately true for the Wallmark diode, but again, signal loss is appreciable since the major part of the generated photocurrent leaks away before reaching the lateral contact

C. Tetra-lateral Diode (Dirichlet Conditions at all Boundaries. Fig. 4)

Assuming lateral contacts along all sides, with negligibly small spacing between adjacent contacts, the current I_{x-} may be derived in a similar fashion as for the duolateral detector

$$I_{x-} = I(x=0;l) = \frac{2}{\pi} I_s \sum_{n=-\infty}^{+\infty} \frac{\sin\left[(2n-1)(\pi y_0/l)\right]}{2n-1} \cdot \left[\frac{\sinh\left[a_n(1-x_0/l)\right]}{\sinh(a_n)} - \sum_{m=-\infty}^{+\infty} \frac{m\pi \cdot \sin\left[m\pi(x_0/l)\right]}{a_n^2 + m^2\pi^2} \cdot \exp\left\{-(a_n^2 + m^2\pi^2)\frac{t}{\beta l^2}\right\}\right] \quad (18)$$

where

$$a_n = \{\alpha^2 l^2 + (2n-1)^2\pi^2\}^{1/2}$$

which may be viewed as a weighted sum of duolateral detector responses with increasing surface recombination.

As for the Wallmark diode, the log ratio of output signals per axis appears to be a good measure for power-compensated position assessment, but for the tetralateral detector, this obtains for almost the whole detector area (see Fig. 7). The central position sensitivity of the log ratio quantity varies linearly with αl for $\alpha l \gg 1$, but it is virtually constant for $\alpha l \lesssim 1$, position linearity improves with increasing $\alpha l > 1$, again at the cost of signal loss due to surface recombination Other quantities such as signal

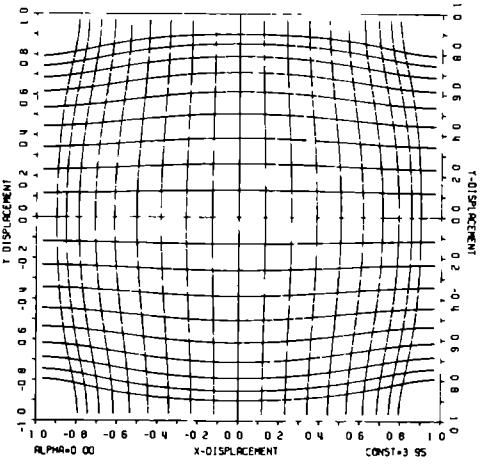


Fig. 7. Dual axis tetralateral diode. Loci for constant values of the log output current ratio per axis log $(X+/X-)$ or log $(Y+/Y-)$, for equidistant fractions of the log ratio CONSTant corresponding to the 90% off-center axial point

current difference per axis, normalized with respect to the signal current sum per axis or total sum, have been found to be less satisfactory.

The steady-state response of the fully reverse-biased, tetralateral diode may also be found by means of conformal mapping, since (2) reduces to the Laplace equation under such circumstances. the inner area of a rectangle may be mapped onto the negative imaginary half plane by means of Weierstrass' $\wp$-function [21]. The rectangular perimeter is mapped onto the real axis, and the boundary condition $U = 0$ may be met by image methods.

LaBaw [8] used image methods directly in two dimensions and found a good agreement between theoretical and experimental results for steady-state current differences per axis. The log ratio position characteristics of a commercial detector[3] appear to agree well with the theoretical curves in Fig 7 [18]. These position characteristics might be rendered more rectilinear by the use of slightly curved lateral contacts; it may be noted in this context that Klein and Bierig [13] reported a linear position characteristic for ratiometric signal processing in a similar, fully reverse-biased configuration with unspecified contact layout and zero loading impedances. An alternative approach might be to exploit the barrel-type distortion exhibited by some optical imaging systems.

Comparing the properties of the three configurations discussed previously, one may conclude that the tetralateral detector ranks between the two other types: its dependency on α is negligible for small values of αl, and

[3] SC/25, United Detector Technology, Inc, Santa Monica, Calif 90405.

TABLE I

Parameter	Symbol	Unit	1	2	3	4	5	6	7	8	9
Reference	—	—	[1]	[3]	[5]	[23]	[7]	[19]	[20]	[24]	[13]
Year	—	—	1930	1957	1960	1960	1962	1965	1965	1970	1974
Material	—	—	Cu–CuO_3	Ge–In	InAs	Si	Ge	Si	Si	Au–Si	Si
Sheet resistivity	ρ_d	Ω·cm		1–2		10–40		100	300	1000	
Sheet thickness	w_d	μm	75	75		13–15			500	170	
Lateral resistance	ρ_d/w_d	kΩ/□		1 3		0 7–3	45		6	60	9
Saturation current density at 300°K	J_d	μA/cm²					500	1		0 4	
Transverse conductance per unit area at zero bias	$\frac{qJ_d}{kT}$	$(kΩ·cm^2)^{-1}$	0 28								
Fall-off parameter	α	cm^{-1}	2 2–5 5		21 1–24 6		9 5	0 25		0 98	0
Contact separation or maximal measuring distance	l	cm	1 3–0 7	0 1	0 14	1–1 3	0 15	0 5–1	1	1	0 6
Normalized fall-off parameter	αl	—	2 8–3 6		3–3 5		1 42	0 13–0 25		0 98	0
Capacitance per unit area at reverse bias	C_d	pF/cm²							3 10^4		42
	U_d	V				0			0		100
Elementary time constant at same bias	βl^2	μs				50–70[a]			200		0 14

[a] It is assumed that the reported values of 5–7 μs correspond with an equivalent single-pole *RC*-filter time-constant of approximately 0 1 βl^2 (see [13])

the position characteristics are rather linear for the log ratio quantity An obvious disadvantage with respect to the dual-axis duolateral configuration of Fig 6 is the non-uniform resolution over the detector surface, for constant light spot power and detector noise currents the signals available from two opposite contacts decrease when the light spot moves out of the center towards either of the remaining contacts

IV TEMPERATURE DEPENDENCY, TRANSIENT RESPONSE, AND NOISE PROPERTIES

Temperature dependency generally increases with α, since the diode reverse saturation current density J_d in (2) is highly sensitive to temperature variations At room temperature, J_d for typical silicon diodes may approximately double per 5°C, this entails a doubling per 10° C of the central position sensitivity for log ratio signals and large αl in the configurations discussed previously, for $\alpha l < 1$, the Wallmark diode becomes even more temperature dependent if the log ratio quantity is used, while the duolateral and tetralateral diodes are quite stable

The impulsive responses of the duolateral and tetralateral detectors have been compared, they may be obtained by analytic differentiation of the step responses with respect to time They appear to consist of a single and slightly delayed impulse which has a fast rising and slow decaying character, at least for moderate values of $\alpha l \lesssim 3$ and not too small distances to the output contact under consideration For decreasing distance, the response approximates a true and delayless Dirac function The tetralateral detector appears to be somewhat faster than the duolateral version as long as surface recombination is of minor influence Although the current *amplitude* of the tetralateral detector varies when the light spot is displaced in parallel with the contact under consideration, the *shape* of the impulse remains virtually unaltered Connors [9] and Klein and Bieng [13] have shown that the step response of the duolateral diode with zero loading impedances in the fully reverse biased mode settles in at most approximately half the elementary time constant βl^2, depending on the distance to the lateral contact This implies practical settling times of 70 ns–100 μS for the parameter values presented in Section V

The fully reverse-biased diode will be faster than its unbiased counterpart since the depletion layer in the former is wider, resulting in a lower capacitance However, if β is considered equal in both modes, settling time decreases with increasing αl for all configurations, at the cost of signal loss due to surface recombination It may be noted in this context, that position resolution in the fully reverse-biased duolateral detector varies linearly with $(C_d l')^{1/2}$, with l' the contact length, as has been derived by Radeka [22] for the case of optimal filtering, this is related to the fact that the Johnson noise in the lateral resistance layer tends to be the dominating noise source in contemporary (Si) detectors [10], [13] The noise properties of lateral detectors are, therefore, inferior to those of ordinary photodiodes which may be shot-noise limited on the other hand, the Johnson noise will hardly be affected by the application of a reverse bias, and it may even decrease if w_d is reduced substantially by the increased depletion layer width

V PRACTICAL PARAMETER VALUES

The literature on the lateral photo-effect has been scanned in order to obtain some typical values for the parameters in the theoretical analysis of the previous sections The entries in Table I are either reported values or have been computed from them Some authors have investigated the lateral effect in quasi-infinite junctions, in those cases, the contact separation distance l has been replaced with the largest distance reported between a measuring position for the transverse potential difference

or output current and an incident light spot or line of light

Thus values of αl may range between 0 and 3 6, while the elementary time-constant βl^2 may vary between 140 ns and 200 μs, for the literature cited The smallest values obtain for the fully reverse-biased mode

On the basis of such parameter values, one may verify whether the small-signal condition leading to (2) and (4) is a practical one The maximum value of $U(x,y)$ will occur at the position of a uniformly illuminated light patch with radius a, where a and α^{-1} are small in comparison with the distance to the nearest boundary Under such circumstances, $U(x,y)$ at the light spot will hardly be influenced by the boundary conditions discussed in this paper, and Lucovsky's analysis [5] for the unbounded case with radial symmetry becomes applicable For values of $\alpha a \ll 1$, his relation (5A) may be approximated as

$$U(r \leq a) \cong \frac{1}{2\pi} \frac{\rho_d}{w_d} I_s\{\ln (1/\alpha a) + 0\,616\} \qquad (19)$$

where

$$r = \{(x - x_0)^2 + (y - y_0)^2\}^{1/2}$$

$I_s = \pi a^2 J_s$ generated photocurrent

Neumann-type boundary conditions at closer distances will tend to augment $U(x_0,y_0)$ while Dirichlet-type boundary conditions have the opposite effect For the linearity hypothesis to hold, one may (arbitrarily) postulate

$$\left|\frac{qU(x_0,y_0)}{kT}\right| \leq 0\,1 \Rightarrow U(x_0,y_0) \leq 2\,5 \text{ mV}, \qquad 300^\circ\text{K} \qquad (20)$$

Assuming a well-focused light spot with $a = 5$ μm, i e , approximately 6 times the wavelength corresponding with the peak sensitivity of typical Schottky-barrier Si-detectors,[3] and using the parameters of a similar detector described by Carr *et al* [24],[4] the following values have been found

$$\alpha a \cong 5\,10^{-4}, \qquad I_s \lesssim 30 \text{ nA}, \qquad J_s \lesssim 40 \text{ mA/cm}^2 \qquad (21)$$

It is of some interest to compare this rather low value for I_s with the noise current originating in the resistance layer Carr *et al* [24] did not present capacity or bandwidth data for the unbiased mode, but one might tentatively use Vasil'yev's parameter values [20] for the settling time $\frac{1}{2}\beta l^2 = 100$ μs This value is in reasonable agreement with observations on the commercial detector referred to before[3] when allowance is made for the different detector areas Assuming a square duolateral configuration, and, neglecting the dependency of the effective noise resistance on frequency [10] and surface-recombination, a value (slightly optimistic) for the Johnson noise current component is found to be

$$I_n = \left(4kTB \frac{w_d}{\rho_d}\right)^{1/2} \cong 50 \text{ pA}_{\text{rms}} \qquad (22)$$

where

$$B = \frac{\pi}{2\beta l^2} \cong 8 \text{ kHz}$$

is the bandwidth which is compatible with the dominant time constant $\beta l^2/\pi^2$ in (13), see [9], [13]

If one assumes a peak-to-peak value of $7I_n$, resolutions in the order of 1:100 for signals obeying the linearity conditions (20) and (21) would seem to be feasible Noise will increase for faster detectors and lower values of ρ_d/w_d I_n in the fully reverse-biased mode may be found to become 5 nA_{rms} for the parameter values reported by Klein and Bierig [13] Thus, for high bandwidths, not only the detector speed should be high, but the signal level for a required resolution may exceed the linearity conditions operation in the fully reverse-biased mode will reduce the capacitance while preserving the linear dependency of the output currents on the incident light distribution Under such circumstances, background (and biasing) components may be compensated by linear operations, even for high background illumination levels For instance, multiple light source position measurement in motion studies is rendered possible through calculations on output current changes resulting from the use of time-multiplexed LED's [18], [25] which are used as position markers on the moving object

Signal levels which are much higher than the linearity bound (21) may easily occur in practice thus laser beams with an optical power of some mW may elicit photocurrents of a few mA, for typical sensitivities of 0 25 A/W [3]

VI CONCLUSIONS

Lateral or longitudinal photo-effects occur because of the diffusion properties of separated charge carriers along uniformly or nonuniformly irradiated p-n junctions The current diffusion may occur due to one or both of the following phenomena 1) Internal surface recombination in a not fully reverse-biased (saturated) junction, which is a basically nonlinear effect since the diode-equation for current versus transverse potential difference is involved, and 2) external collection of generated charge carriers through finite loading impedances These two types of lateral effect have been described as being completely distinct from each other [26], but it should be realized that both phenomena may occur in a particular diode depending upon external circumstances, such as biasing potential and terminating impedances

A unified analysis of the transient and steady-state responses of rectangular detectors operated in the small-signal (unbiased) and fully reverse-biased modes has been presented, using a linear 2-dimensional diffusion model The steady-state position characteristics for power-com-

[4] The lateral resistance ρ_d/w_d was calculated by means of the reported values of J_d in Fig 3 and α_0 in Section IV of Carr *et al* [24]

pensated output quantities have been discussed in some depth, with emphasis on practical aspects such as position linearity and invariance with device parameter changes. The following conclusions have been drawn.

1) The dual-axis, duolateral detector is most promising because of its inherent linearity and uniform resolution when surface recombination may be neglected.[8]

2) The tetralateral detector is somewhat faster than the single-axis duolateral diode but exhibits a minor nonlinearity and nonuniform resolution for the log ratio of output currents per axis quantity.

3) The unloaded Wallmark configuration appears to have less favorable properties in terms of linearity and temperature dependency.

The limited utility of operation in the small-signal mode has been illustrated with some typical parameter values. For the following reasons, the fully reverse-biased detector with zero loading impedances is superior to configurations in which surface recombination occurs.

1) Signal loss is reduced, permitting lower incident light levels and larger diode dimensions.

2) Response speed and resolution may improve because of decreased capacitance which is a fixed quantity under these circumstances.

3) The output currents are linear functions of the incident light distribution, even for high illumination levels: thus strong background components in the output currents may be compensated by linear operations in the time- or frequency domains prior to ratiometric signal processing for power compensation.

4) Temperature dependency may be negligible: a uniform change in the resistivity of the resistance sheet has no effect upon the power-compensated steady-state position sensitivity (apart from noise effects), since ρ_d is involved through α only, and α may be considered to equal zero in this mode.

VII. ACKNOWLEDGMENT

The author would like to express his gratitude to Dr. E. G. J. Eijkman and Dr. J. H. J. Klijn for critical proof reading of an earlier version of the manuscript, Mrs. D. Dijckhoff-Laenen for the drawings, Dr. D. Cotton for linguistic suggestions, and many others who have contributed to the results reported here.

[8] Ion-implanted silicon photodetectors of this type have very recently been constructed in Wallmark 's laboratory at Chalmers University of Technology, Gothenburg, Sweden; a position linearity to within 0.1-percent FS has been attained for an active area of 24 × 24 mm² at a reverse bias of 10 V [27].

VIII. REFERENCES

[1] W. Schottky, "Ueber den Entstehungsort der Photoelektronen in Kupfer-Kupferoxydul-Photozellen," *Phys. Z.*, Leipzig, vol. 31, pp. 913–925, Nov. 1930.
[2] R. B. Emmons, "The lateral photoeffect," *Solid-State Electron.*, vol. 10, pp. 505–506, May 1967
[3] J. T. Wallmark, "A new semiconductor photocell using lateral photoeffect," *Proc. IRE*, vol. 45, pp. 474–483, Apr. 1957.
[4] A. R. Moore and W. M. Webster, "The effective surface recombination of a germanium surface with a floating barrier," *Proc. IRE*, vol. 43, pp. 427–435, Apr. 1955.
[5] G. Lucovsky, "Photo-effects in nonuniformly irradiated p-n junctions," *J. Appl. Phys.*, vol. 31, pp. 1088–1095, June 1960.
[6] D. A. Allen, "An analysis of the radiation tracking transducer," *IRE Trans. Electron Devices*, vol. ED-9, pp. 411–416, Sept 1962.
[7] I. I. Taubkin and A. L. Frimer, "Calculation of longitudinal-effect photocells," *Sov. Radio Eng.*, vol. 7, pp 1120–1129, 1962.
[8] K. B. LaBaw, "Theoretical model of the lateral photodiode," NWC-IDP 3161, Naval Air Facility, Aviation Ordn. Dept., China Lake, Calif., 1970
[9] W. P. Connors, "Lateral photodetector operating in the fully reverse-biased mode," *IEEE Trans. Electron Devices*, vol. ED-18, pp. 591–596, Aug 1971
[10] R. B. Owen and M. L. Awcock, "One- and two-dimensional position sensing semiconductor detectors," *IEEE Trans. Nucl. Sci.*, vol. 15, pp. 290–303, June 1968
[11] S. Kalbitser and W. Stumpfl, "A nomogram for the design of position sensitive silicon detectors," *Nucl. Instrum. Methods*, vol. 77, pp. 300–302, Jan. 1070.
[12] J. R. Gigante, "Si(Li) position sensitive detector," *Nucl. Instrum. Meth.*, vol. 111, pp 345–354, Aug /Sept. 1973.
[13] C. A Klein and R. W Bierig, "Pulse-response characteristics of position-sensitive photodetectors," *IEEE Trans. Electron Devices*, vol. ED-21, pp. 532–537, Aug 1974.
[14] Y. Marfaing, "Propriétés électriques des jonctions d'antimoniure d'indium de grande surface," *Solid-State Electron.*, vol 7, pp 1–16, Jan 1964
[15] H. S Carslaw and J. C Jaeger, *Conduction of Heat in Solids*, 2nd ed. Oxford, England: Clarendon Press, 1959
[16] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*. New York: McGraw-Hill, 1961.
[17] L. B W. Jolly, Ed., *Summation of Series*. New York: Dover, 1962.
[18] H. J. Woltring, "New possibilities for human motion studies by real-time light spot position measurement," *Biotelemetry*, vol. 1, pp. 132–146, 1974.
[19] L. K. Buzanova, A. M. Vasil'yev, A. Ya Gliberman, and A. P. Landsman, "Photocells exhibiting the longitudinal photoeffect," *Sov. Radio Eng.*, vol. 10, pp. 109–116, 1965.
[20] A. M. Vasil'yev, "Transient processes in photocells with the longitudinal photoeffect," *Sov. Radio Eng.*, vol. 10, pp. 1575–1579, 1965.
[21] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1965.
[22] V. Radeka, "Signal, noise, and resolution in position-sensitive detectors," *IEEE Trans. Nucl. Sci.*, vol. 21, pp. 51–64, Feb. 1974.
[23] D. Allen, I. Weinman, and J. Winslow, "Radiation tracking transducer," *IRE Trans. Instrumentation*, vol. I-9, pp. 336–341, Dec. 1960.
[24] Th. G. Carr, J. C. Richmond, and R. G. Wagner, "Position-sensitive Schottky-barrier photodiodes: Time-dependent signals and background saturation effects," *IEEE Trans. Electron Devices*, vol. ED-17, pp. 507–513, July 1970.
[25] L.-E. Lindholm and K. E. T. Oeberg, "An optoelectronic instrument for remote on-line movement monitoring," in *Proc. 2nd Int. Symp. Biotelemetry*, P. A. Neukomm, Ed. Basel, Switzerland: Karger, 1974.
[26] B O. Kelly and R. I. Nemhauser, "Techniques for using the position sensitivity of silicon photodetectors to provide remote machine control," 21st Annu IEEE Machine Tool Conf., Hartford, Conn., 1973.
[27] G. Pettersson, private communication.

Steady-state loci for constant output signals for the square Wallmark photodiode with unloaded mid-lateral point contacts

The curves on the following pages display loci for constant output signals such as log ratio of transverse potentials per axis or potential difference per axis, for equidistant fractions of a value CONSTant corresponding to the signal value at the off-centre axial point. The relative increment is 1/9.

ALPHA denotes the product of Lucovsky's lateral fall-off parameter α and the diode side length l. ALPHA assumes the values 0.1, 1 and 10.

X+ Transverse potential at (+1, 0)

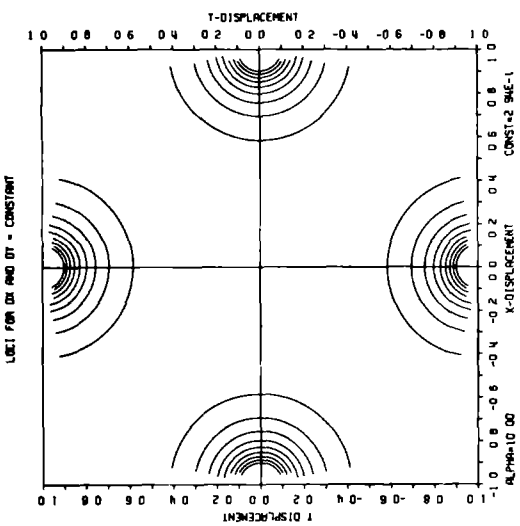
X− Transverse potential at (−1, 0)

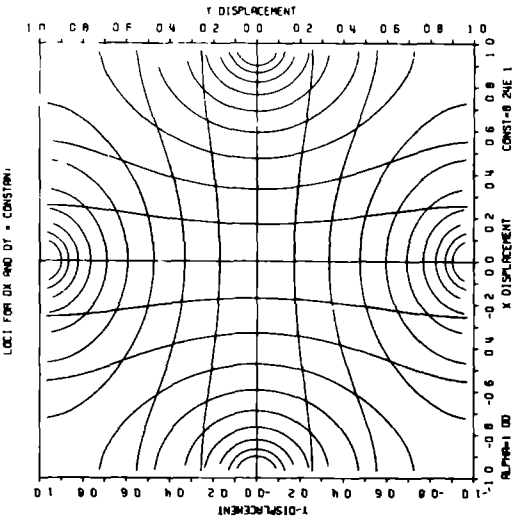
Y+ Transverse potential at (0, +1)

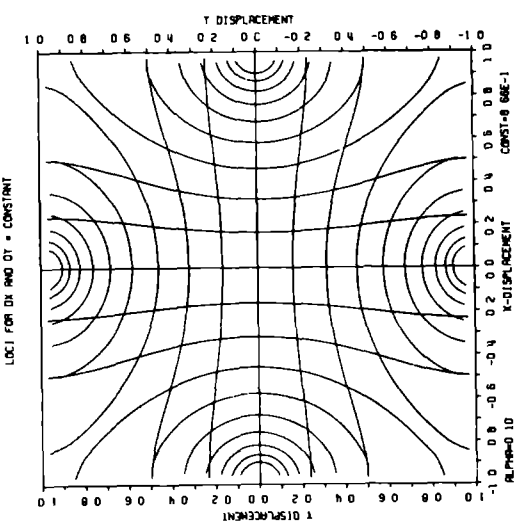
Y− Transverse potential at (0, −1)

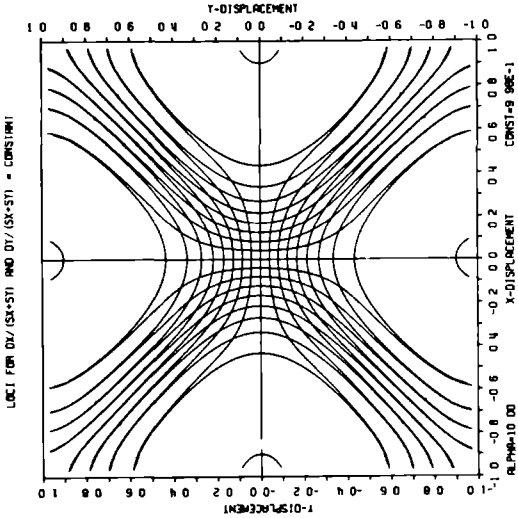
DX = X+ − X− DY = Y+ − Y−

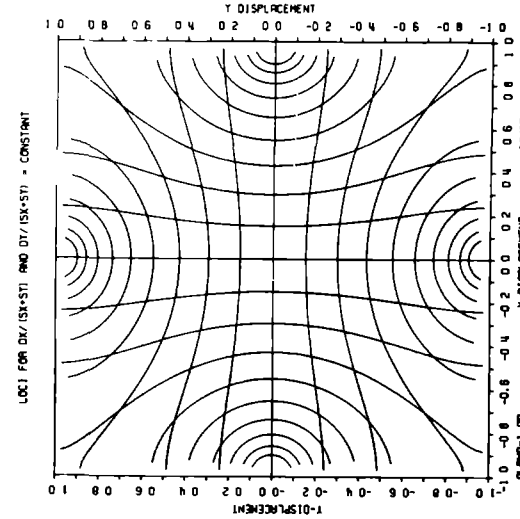
SX = X+ + X− SY = Y+ + Y−

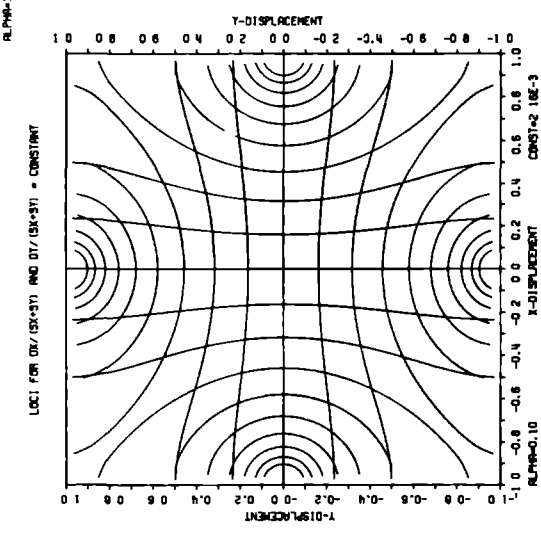

T-DISPLACEMENT
LOCI FOR DX AND DY = CONSTANT
X-DISPLACEMENT
T DISPLACEMENT
ALPHA=10 00
CONST=2 BAE-1

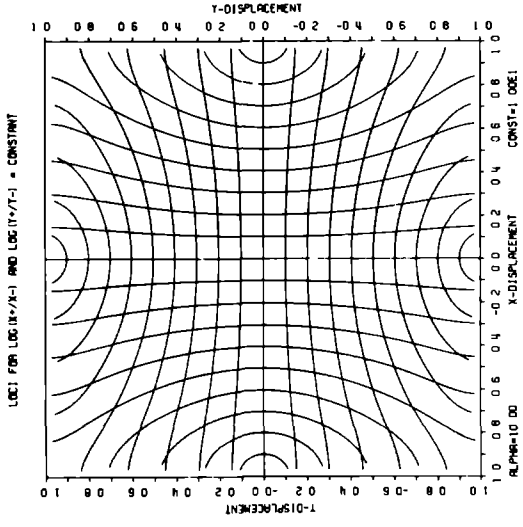

Y DISPLACEMENT
LOCI FOR DX AND DY = CONSTANT
X DISPLACEMENT
T-DISPLACEMENT
ALPHA=1 00
CONST=0 2AE 1

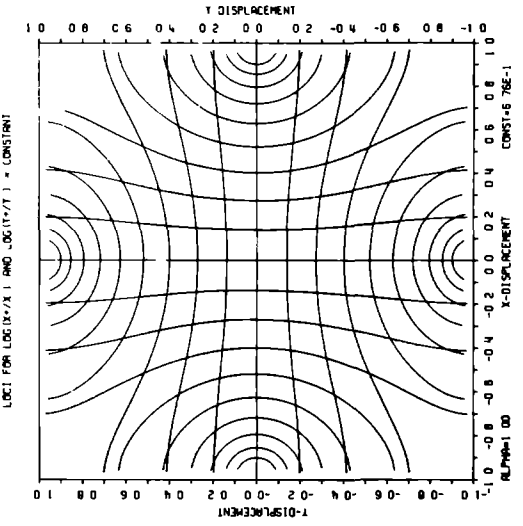

Y DISPLACEMENT
LOCI FOR DX AND DY = CONSTANT
X-DISPLACEMENT
T DISPLACEMENT
ALPHA=0 10
CONST=0 6AE-1

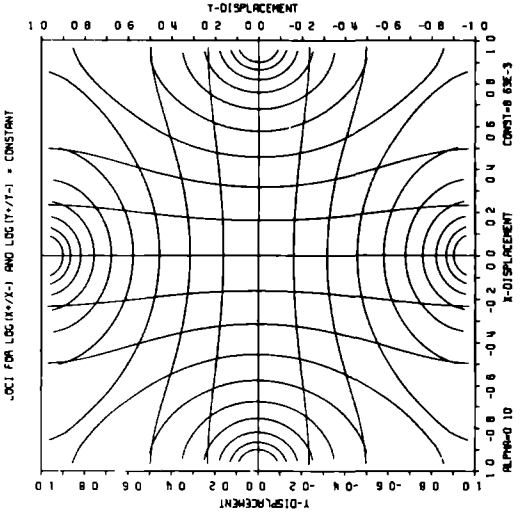

LOCI FOR DX/(SX*ST) AND DY/(SX*ST) = CONSTANT
Y-DISPLACEMENT
X-DISPLACEMENT
T-DISPLACEMENT
ALPHA=10 00
CONST=9 99E-1


LOCI FOR DX/(SX+ST) AND DY/(SX+ST) = CONSTANT
Y DISPLACEMENT
X-DISPLACEMENT
T-DISPLACEMENT
ALPHA=1 00
CONST=1 00E-1


LOCI FOR DX/(SX+ST) AND DY/(SX+ST) = CONSTANT
Y-DISPLACEMENT
X-DISPLACEMENT
T-DISPLACEMENT
ALPHA=0.10
CONST=2 10E-3


LOCI FOR LOG(X+/X-) AND LOG(Y+/Y-) = CONSTANT
Y-DISPLACEMENT
X-DISPLACEMENT
T-DISPLACEMENT
ALPHA=10 00
CONST=1 00E1


LOCI FOR LOG(X+/X-) AND LOG(Y+/Y-) = CONSTANT
Y DISPLACEMENT
X-DISPLACEMENT
T-DISPLACEMENT
ALPHA=1 00
CONST=6 78E-1


LOCI FOR LOG(X+/X-) AND LOG(Y+/Y-) = CONSTANT
Y-DISPLACEMENT
X-DISPLACEMENT
T-DISPLACEMENT
ALPHA=0 10
CONST=8 63E-3

Steady-state loci for constant output signals for the tetra-lateral photodiode with zero loading impedances

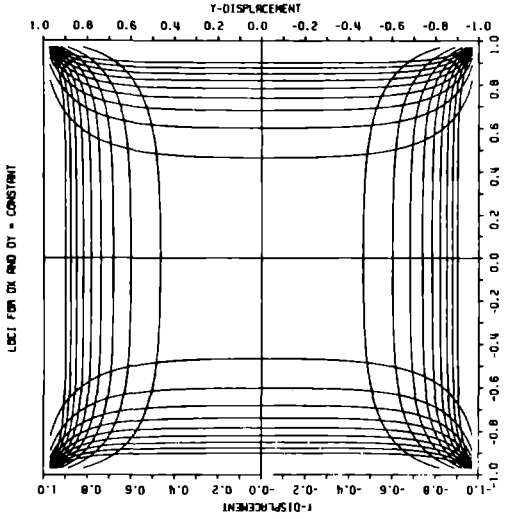
The curves on the following pages display loci for constant output signals such as log output currents ratio per axis or current difference per axis, for equidistant fractions of a value CONSTant corresponding to the signal value at the 90% off-centre axial point. The relative increment is 1/9.

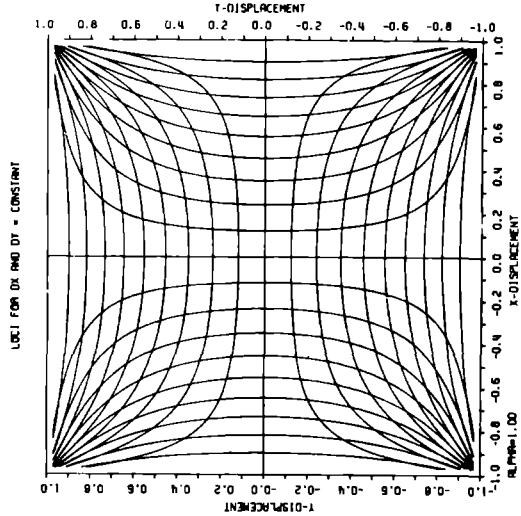
ALPHA denotes the product of Lucovsky's lateral fall-off parameter α and the diode side length l. ALPHA assumes the values 0, 1 and 10.

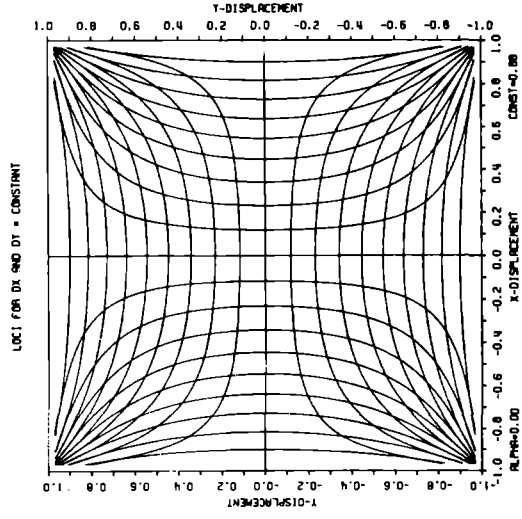
X+	Output current at positive X-contact
X−	Output current at negative X-contact
Y+	Output current at positive Y-contact
Y−	Output current at negative Y-contact

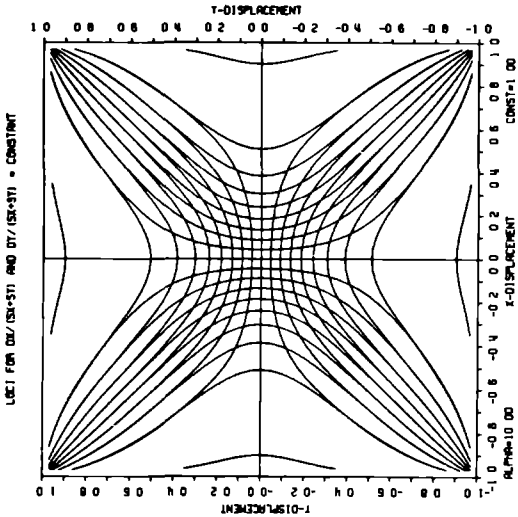
DX = X+ − X− DY = Y+ − Y−

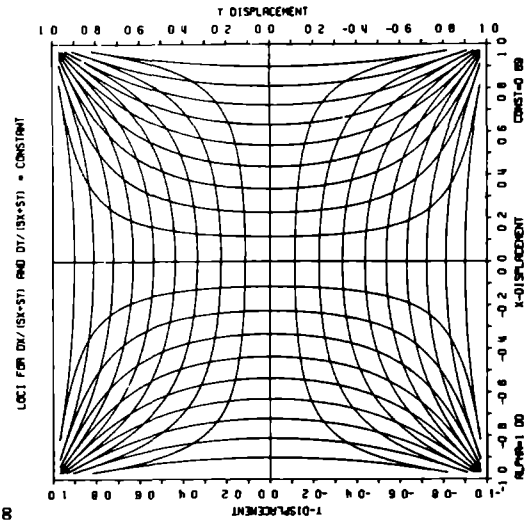
SX = X+ + X− SY = Y+ + Y−

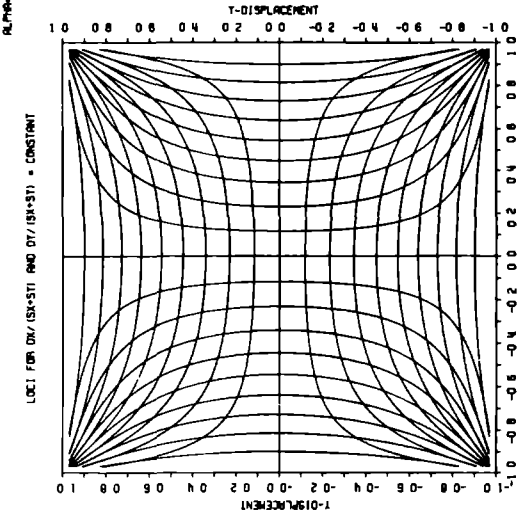

Y-DISPLACEMENT
1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0
LOCI FOR DX AND DY = CONSTANT
X-DISPLACEMENT
CONST=0.81
ALPHA=10.00
1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0
Y-DISPLACEMENT

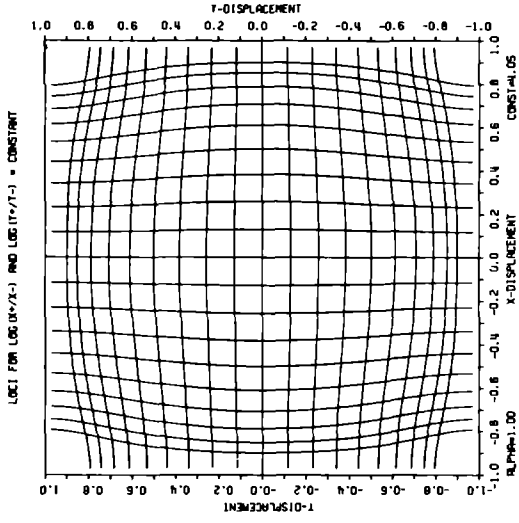

Y-DISPLACEMENT
1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0
LOCI FOR DX AND DY = CONSTANT
X-DISPLACEMENT
ALPHA=1.00
1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0
Y-DISPLACEMENT

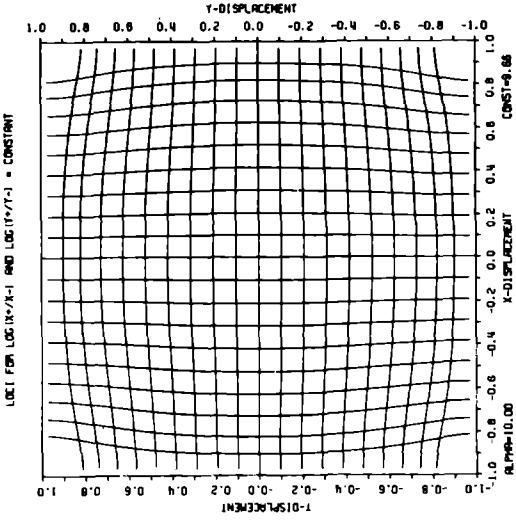

Y-DISPLACEMENT
1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0
LOCI FOR DX AND DY = CONSTANT
X-DISPLACEMENT
CONST=0.00
ALPHA=0.00
1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0
Y-DISPLACEMENT

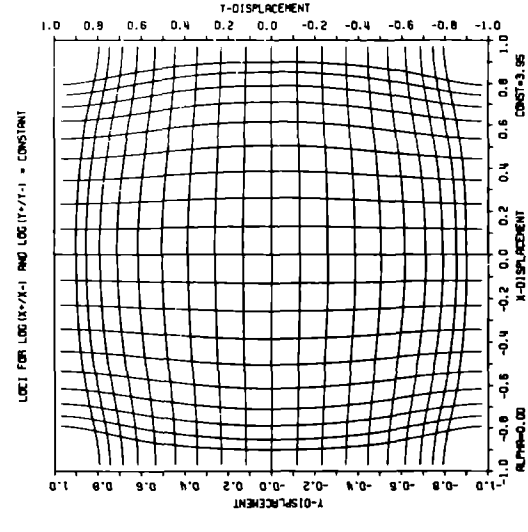

T-DISPLACEMENT
LOCI FOR DX/(SX+ST) AND DY/(SX+ST) = CONSTANT
X-DISPLACEMENT
T-DISPLACEMENT
ALPHA=1.00
CONST=1.00


T DISPLACEMENT
LOCI FOR DX/(SX+ST) AND DY/(SX+ST) = CONSTANT
X-DISPLACEMENT
T-DISPLACEMENT
ALPHA=1.00
CONST=0.00


T-DISPLACEMENT
LOCI FOR DX/(SX+ST) AND DY/(SX+ST) = CONSTANT
X-DISPLACEMENT
T-DISPLACEMENT
ALPHA=0.00
CONST=0.00


Y-DISPLACEMENT
LOCI FOR LOG(X'/X-) AND LOG(Y'/Y-) = CONSTANT
X-DISPLACEMENT
CONST=4.05
ALPHA=1.00
T-DISPLACEMENT


Y-DISPLACEMENT
LOCI FOR LOG(X'/X-) AND LOG(Y'/Y-) = CONSTANT
X-DISPLACEMENT
CONST=9.88
ALPHA=10.00
T-DISPLACEMENT


Y-DISPLACEMENT
LOCI FOR LOG(X'/X-) AND LOG(Y'/Y-) = CONSTANT
X-DISPLACEMENT
CONST=3.95
ALPHA=0.00
Y-DISPLACEMENT

Steady-state output signals of the fully reverse-biased tetra-lateral photodiode with zero loading impedances ($\alpha = 0$)

The curves on the following pages display the output signal dependancy on variation of a small light spot of constant power along the x-axis, for constant values of the y-coordinate.

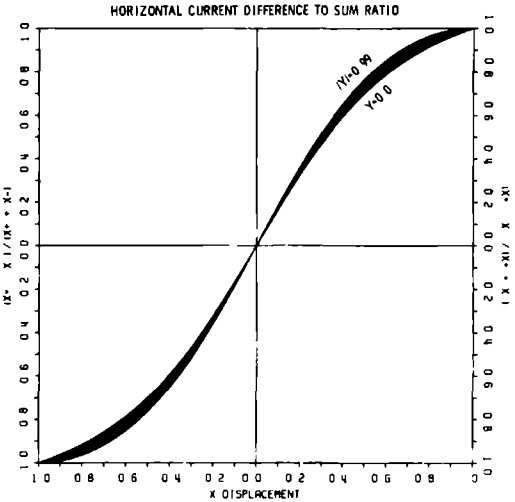
The values /y/ = 0.0(0.1)0.9 and 0.99 have been chosen. For reasons of symmetry, the curves for y = +c and y = −c coincide in a few cases.

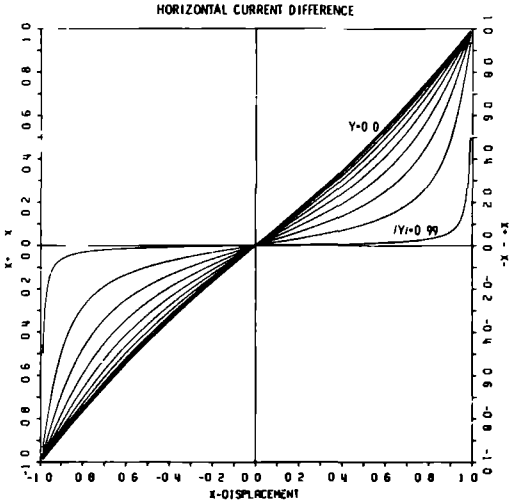
X+ Output current at positive X-contact

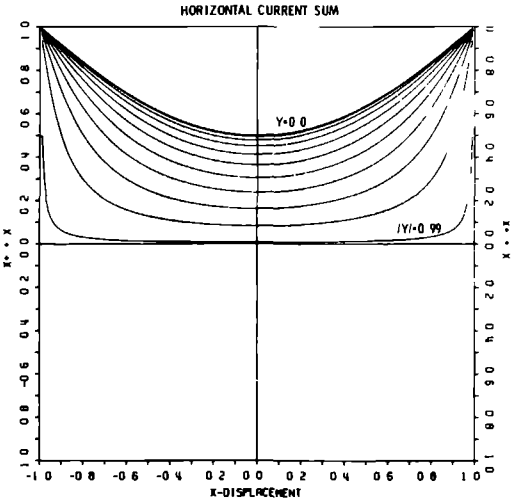
X− Output current at negative X-contact

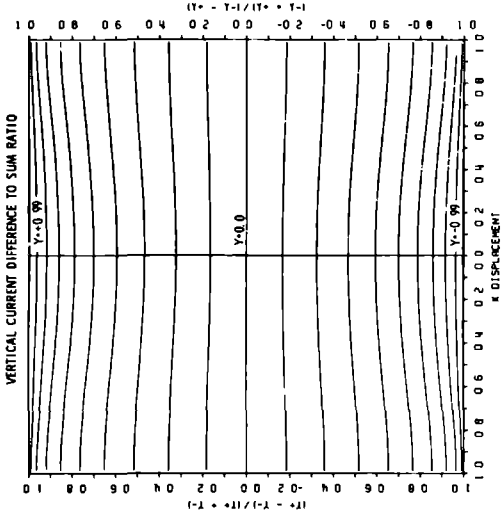
Y+ Output current at positive Y-contact

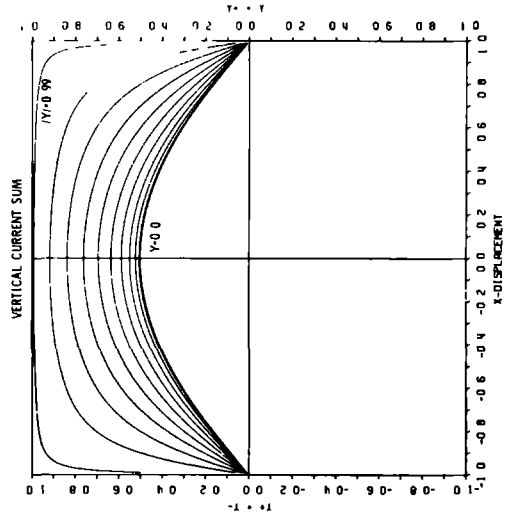
Y− Output current at negative Y-contact

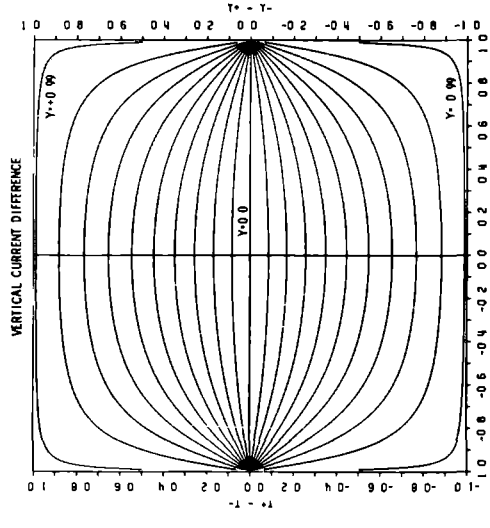

HORIZONTAL CURRENT DIFFERENCE TO SUM RATIO
/Y/=0 99
Y=0 0
X DISPLACEMENT

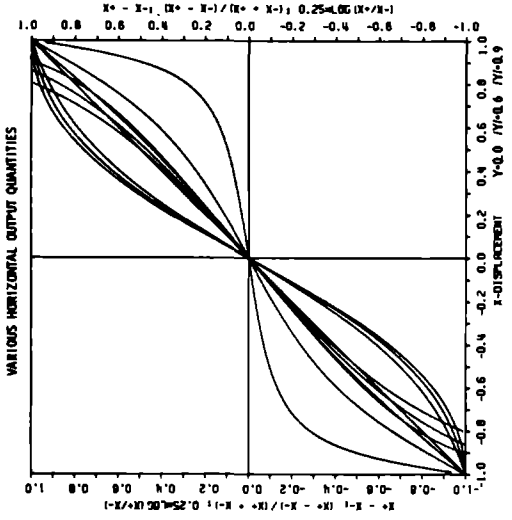

HORIZONTAL CURRENT DIFFERENCE
Y=0 0
/Y/=0 99
X-DISPLACEMENT

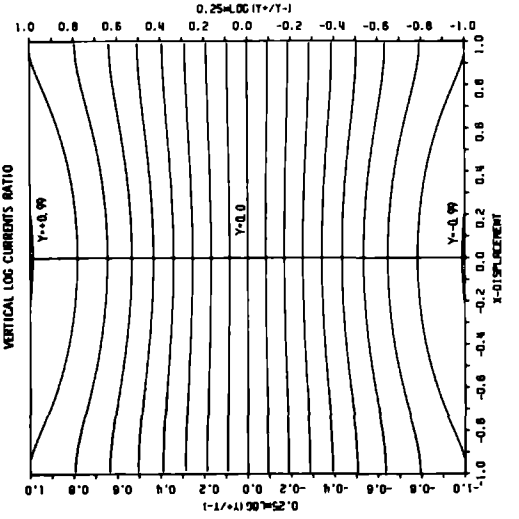

HORIZONTAL CURRENT SUM
Y=0 0
/Y/=0 99
X-DISPLACEMENT

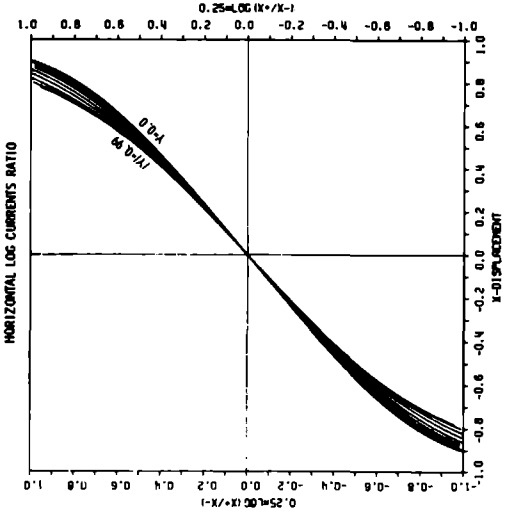

VERTICAL CURRENT DIFFERENCE TO SUM RATIO
Y=0.99
Y=0.0
Y=-0.99
X DISPLACEMENT


VERTICAL CURRENT SUM
Y=0.99
Y=0.0
X-DISPLACEMENT


VERTICAL CURRENT DIFFERENCE
Y=0.99
Y=0.0
Y=-0.99
X-DISPLACEMENT


VARIOUS HORIZONTAL OUTPUT QUANTITIES
X-DISPLACEMENT
Y=0.0
Y=0.6
Y=0.9


VERTICAL LOG CURRENTS RATIO
0.25*LOG(Y+/Y-)
X-DISPLACEMENT
Y=0.0
Y=0.99
Y=-0.99


HORIZONTAL LOG CURRENTS RATIO
0.25*LOG(X+/X-)
X-DISPLACEMENT
Y=0.0
Y=0.99

Chapter II

Data Processing Aspects

Biotelemetry
2: 169-196 (1975)

Punkten är den minsta linjen.

Fallstaff, Fakir: En var sin egen professor - om geometri.

2.1.1

Calibration and Measurement in 3-Dimensional Monitoring of Human Motion by Optoelectronic Means

I. Preliminaries and Theoretical Aspects

H. J. WOLTRING

Laboratory of Psychology, University of Nijmegen, Nijmegen

Key Words. Motion study · Biomechanics · Displacement sensing · Transducers Software · Measuring equipment · Photogrammetry · Light telemetry

Abstract. By combining the observations from a number of independent direction sensors, the spatial positions of one or more targets may be reconstructed. The theory of calibration and measurement in a computerized, interactive environment is discussed in the terminology of analytical photogrammetry, for applications in the field of biomechanics. The discussion is focussed upon optoelectronic systems for movement monitoring in real-time rather than upon photography as is the conventional photogrammetric implementation.

Introduction

Three-dimensional (3-D) measurement of shapes, positions and movements is a recurring problem in a variety of applications including surveying, cartography, civil engineering, archeology and ballistics. In the life sciences, such applications include tomography, prosthesiology, cytology and biomechanics. An important class of measuring methods is based on observing a set of 2-dimensional (2-D) projections of the structure under investigation, from which a 3-D estimate may be subsequently constructed. Traditional registration devices are of a photographic or cinematographic nature [15]; more recently, optoelectronic methods have been

increasingly applied, especially for dynamic problems as occur in ballistics and biomechanics [22, 24].

Projective data may be based on parallel radiation incident on a transducer, as in the case of transmission pictures in radiology. If the orientations and positions during exposure of a sufficient set of observed images are known, the 3-D density distribution of the original structure may be reconstructed by linear operations; various algorithms are in use for this purpose [2, 5]. Alternatively, projective data may be obtained from convergent radiation passing through a lens system; unlike parallel projection, this is a nonlinear transformation of the original structure rendering the reconstruction problem more complicated. However, data acquisition is more flexible since variations in the area to be covered may be accounted for by distance or focus variation.

3-D reconstruction from convergent projections is the central issue in the science of *photogrammetry* [10, 19, 23]; a large variety of instrumental and analytical tools have been developed in this field. Popular implementations for *instrumental photogrammetry* are stereoviewers which display slightly different photographs to each eye, and anaglyphs which are pictures in different colours to be viewed through spectacles containing corresponding colour filters.

Analytical photogrammetry is based on formal analysis in terms of mathematical models; it has its origin in the development of projective geometry as initiated by the theorems of the Renaissance painters. Again for a sufficient set of projective observations and knowledge of camera parameters such as position, attitude, focal length and image distortion, the original 3-D structure may be reconstructed. Since the calculations involved are quite time-consuming, analytical photogrammetry has been of mainly theoretical importance up to the time that modern computational equipment became available; many recent results have been obtained in ballistic research and aerial surveying.

In the present paper, a number of procedures for 3-D measurement of human motion will be discussed in photogrammetric terms. As such, it is a rather formal analysis which pertains equally to photographic, cinematographic or optoelectronic camera systems.

The software aspects of the implementation of these procedures with minicomputers will be discussed to some extent. The high data rates resulting from monitoring a number of body markers with TV-cameras or other electronic equipment [22, 24] at typical sampling intervals of 1–20 ms virtually necessitate computer support. Minicomputers are

quite suitable for such purposes; they tend to be available on a real-time, on-line basis allowing motion data to be reduced and analyzed during rather than after an observation period – although this feature has its limitations, as will become clear later.

Both system calibration and target measurement will be discussed, with emphasis on real-time movement monitoring and calibration flexibility. Whereas in many photogrammetric applications, calibration and measurement may or even must be performed simultaneously, by entering known positional information on observed *control points*, separation of calibration and measurement is possible in close-range, real-time monitoring. Thus, flying height in aerial photogrammetry must be obtained from observed photographic data *a posteriori* since no other information of sufficient accuracy may be available, whereas the observing cameras in biomechanical research may be calibrated *a priori,* if they have fixed position and attitude during measurement. Constancy of camera parameters during measurement is assumed throughout this paper, with certain consequences on observation area dimensions; the alternative approach of cameras which vary their relative positions by mechanically tracking a moving subject is not considered.

The choice of camera parameters depends on the application, e.g., in gait analysis, one will choose certain combinations of intercamera distance, camera orientations with respect to each other, and observation distance which differ from those adopted in, for instance, arm movement studies. In addition, it is important that the observed directions intersect at sufficiently large *parallactic angles,* in order to attain an acceptable reconstruction accuracy as discussed later.

An automatic calibration method will render such parameter freedom of practical use, and the common photogrammetric approach is to estimate parameters by observing a number of control points whose locations are fully or partially known in terms of some predefined, operational coordinate system. By interfacing a movement-monitoring system to a computer, procedures of this kind can be realized. Prior to a measurement series, a calibration object with a number of control points is displayed and subsequently removed. For practical reasons, it is advantageous to employ flat objects: regularly perforated metal plates are commonly available, and one may simply illuminate such *control planes* from behind. Alternatively, some systems are not capable of simultaneously discriminating more than one light point [24], and one may have to display a number of separately controlled light sources at regular intervals.

Some flexibility can in principle be attained by mounting two cameras on a rigid bar with fixed intercamera distance or *base* length and with fixed orientation angles with respect to each other, usually in parallel; such *stereometric cameras* are quite common in close-range photogrammetry. For known base and focal lengths, the original 3-D object may be reconstructed in terms of a coordinate system defined by the stereometric camera. Usually, one will prefer to describe the object in other terms, and calibration of the stereometric camera with reference to an operationally defined coordinate system remains necessary. A disadvantage of these cameras is that the parallactic angles are usually rather small, with consequent loss in depth accuracy.

A traditional distinction exists between *external* and *internal* camera parameters. The former concern externally manipulable parameters such as camera position and attitude, and sometimes focal length, while the latter concern position of the image plane with respect to the optical axis, and image distortion. Estimation of these parameters from a set of control points is commonly denoted as space *resection* and camera *calibration*, respectively; the opposite problem of 3-D position reconstruction of observed targets is called space *intersection* (of light rays).

Calibration methods in analytical photogrammetry usually separate the estimation of both parameter types: internal parameters are estimated in a laboratory environment, and external parameters are obtained by control point observations 'in the field'. Unfortunately, the distinction is somewhat artificial since allegedly internal parameters may depend on external parameters; an example is the influence of observation distance and lens aperture on image distortion. Moreover, drift in electronic devices or errors in positioning photographs may cause deviations from previously assessed internal parameter values. A current trend is therefore to estimate all parameters 'in the field' [11], and this approach has been adopted in this paper. Consequently, the distinction between both parameter classes looses much of its meaning, and the term calibration will be used in a general sense unless otherwise indicated. Correlation between error sources or errors in interfacing components are revealed in this approach; they are not accounted for in piecewise calibration procedures.

The basic method is to postulate a mathematical model for both the true projective relations and for the deviations from this ideal, with unknown parameters to be estimated in some optimal sense. The two aspects are sometimes called first order and second order theory, respectively.

The control point distribution in calibration is of paramount importance. A typical problem is the confounding of unknown focal length and observation distance for a single control plane parallel to the image plane [11, 16]. These parameters become separable in the case of oblique projection, but they remain quite error-prone because of their high intercorrelations. As a consequence, position and attitude errors may compensate each other.

The *geometrical strength* of the combination of control point distribution and mathematical model may be improved by employing a 3-D calibration object [11] such as a number of control planes in parallel. This causes some new problems, since the relative attitudes and positions of these control planes must be known. In a fully general approach, one may have to estimate these control parameters simultaneously, in the sense of intervening variables.

The resulting equations are usually nonlinear and not explicitly solvable, both in calibration and measurement. Iterative or implicit estimation techniques are employed under such circumstances, a process which is commonly denoted as (camera or target) *adjustment*. In the case of calibration, one has in essence an imaginary camera which exists as a software procedure; the parameters of this modelled camera are adjusted in such a way as to minimize, for instance, the discrepancy between observed and modelled camera outputs.

Some theoretical and practical considerations on such estimation procedures are presented in the following sections. After a discussion on photogrammetric fundamentals [19] and on systematic errors, the type and solution of equations for calibration and measurement are considered. It will become apparent that requirements of calculation speed determine the type of equations and optimization criteria to be used, with consequences on attainable accuracy. Subsequently, criteria for choosing suitable control point distributions and camera parameters are presented, and some experimental results are finally given as an illustration of the previous theory. The latter have been obtained with a preliminary version of a commercial, optoelectronic system for 3-D movement monitoring 'SELSPOT', an acronym for SELective light SPOT recognition [9, 14].

This system is based on photodiodes exploiting the *lateral photoeffect* [24, 25] to determine the mean position of an incident light distribution; it allows the observation of up to 30 light-emitting diodes (LEDs) which may be affixed to the human body. The system has a bandwidth of over 100 Hz and spatial resolution of up to 10 bits per axis, depending on

the intercepted light power. More extensive results will be presented in part II of this paper [26].

Photogrammetric Fundamentals (First Order Theory)

Spatial measurements imply the use of coordinate systems in terms of which points or objects are characterized. Basically, analytical photogrammetry consists of the mathematical description of relations between points and lines in two different coordinate systems: (1) the *exterior* or *object* coordinate system, in which observed events are to be reconstructed (capital letters); (2) the *interior* or *image* coordinate system, in which the observations are expressed (small letters).

The relation between the coordinates $\mathbf{X}_P$ and $\mathbf{x}_P$ of a point P in the two systems is:

$$\mathbf{x}_P - \mathbf{x}_C = M\,(\mathbf{X}_P - \mathbf{X}_C), \tag{1}$$

with C some other point with coordinates $\mathbf{X}_C$ and $\mathbf{x}_C$, respectively, and M an *orientation matrix* with respect to the external coordinate system.

Object Coordinate System

The definition of the external coordinate system varies with the application: in land surveying one may adopt a Cartesian coordinate system defined by the vertical, by the north direction and by their common perpendicular; in space navigation one may employ a stellar or solar coordinate system, and in cartography the geocentric system. For close-range applications such as in biomechanics, a simple right-hand Cartesian system will mostly suffice. From data in this object system, one may subsequently wish to derive information in terms of a corporal coordinate system [15, 20] which moves with respect to the cameras.

Image Coordinate System

Usually, the image plane serves to identify the xy-plane, and the origin is defined with respect to externally determinable entities such as photograph edges ('fiducial marks') or full-scale output signals (fig. 1). The optical axis is considered to be normal to the image plane and is, therefore, parallel to the z-axis. The point of intersection of image plane and optical axis is called the *principal point* o, with coordinates $(x_o, y_o, 0)^T$. Ideally, the principal point coincides with the image origin, but alignment errors

may cause small deviations. Unless otherwise stated, both points will be assumed to coincide.

Orientation Matrices

Orientation[1] matrices consist of 9 correlated elements which may be defined in terms of various rotations. One such a combination adopted in this paper is the successive rotation about the X-axis with an angle ω, about the Y-axis with an angle φ, and about the Z-axis with an angle $\varkappa$. Thus, M becomes

$$M(\varphi) = K(\kappa) \cdot \Phi(\varphi) \cdot \Omega(\omega) \tag{2}$$

with $\varphi \triangleq (\omega, \varphi, \kappa)^T$ (^{T} denotes a transposed vector or matrix), and

$$K(\kappa) = \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix}, \Phi(\varphi) = \begin{bmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{bmatrix}, \Omega(\omega) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{bmatrix}.$$

For pure orientation matrices M, the inverse M^{-1} is equal to the transpose M^T.

Relation between Object and Image Points: Projective Equations

In imagery by convergent projections, all incident light rays are assumed to intersect at a common point on the optical axis: the *perspective centre* C with image coordinates $(x_o, y_o, c)^T$. For the ideal pinhole camera, the pinhole itself is the perspective centre; for practical, lens-equipped cameras, the perspective centre is somewhere inside the optical system.

If the object is sufficiently far away from the camera, e.g. in the case of aerospace photogrammetry, c equals the focal length of the camera; under close-range circumstances, c will vary slightly with the object-to-camera distance in accordance with the well-known lens formula relating object and image distance to focal length. For a fixed camera, close-range distance variations will cause both image blurring and distortion. Blurring affects image detail, and thus positional accuracy in a fairly unpredictable fashion; distortion causes systematic image deformation which may be calibrated. In practice, image sharpness will be sufficient within a given

[1] In photogrammetric practice, the notion *outer orientation parameters* refers to the 6 external parameters of rotation (φ) and translation (X_c). In order to avoid confusion, the terms camera *attitude* and camera *position* (station) are preferred. Similarly, the notion *inner orientation parameters* for focal length, principal point and image distortion parameters is avoided.

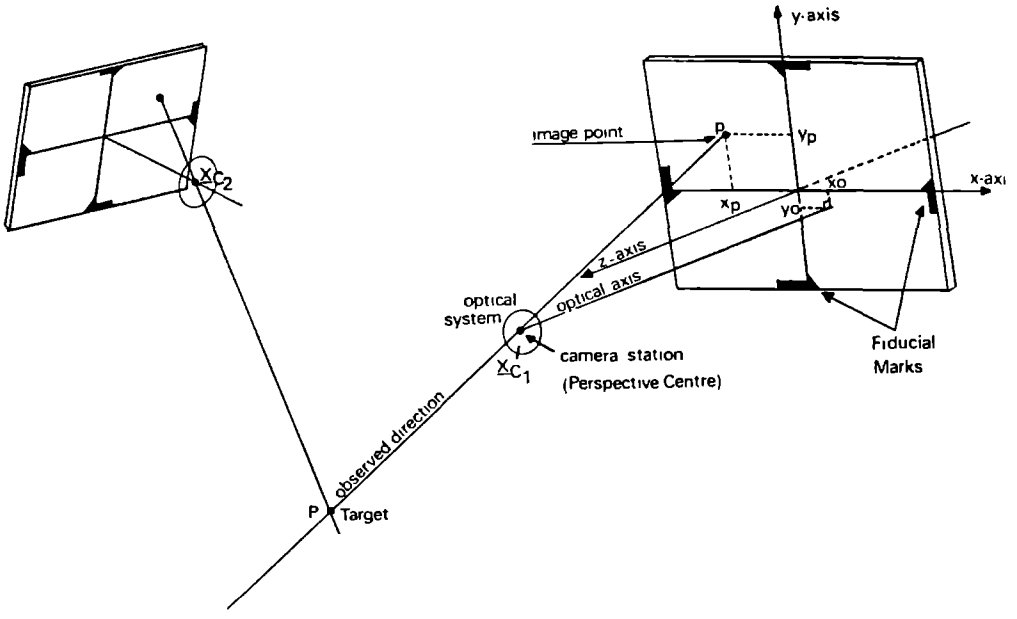


Fig. 1. Relation between external and internal quantities in space intersection.

and predeterminable distance range, and c is assigned the value of what is called the *camera constant* or *calibrated focal length* of the particular set-up, involving also an equilibration of image distortion (see below).

Since light rays may be assumed straight in close-range circumstances, a simple relation exists between the coordinates of an object point P, of its image p and of the perspective centre C

$$\begin{aligned} X_P - X_C &= k \cdot M^T (x_p - x_c) \\ &= k \cdot M^T (x_p, y_p, -c)^T \text{ (assuming } x_o = y_o = 0), \end{aligned} \tag{3}$$

with k a proportionality factor.

These *projective equations* formalize that P, p and C are on a common straight line (fig. 1); they may perhaps be viewed as the most important photogrammetric equations.

The various notions introduced until now are illustrated in figure 1, for ideal cameras. This idealization does not obtain in practice since lenses exhibit distortion and have finite dimensions. As a consequence, relation (3) may be used only after correction of raw observation data. Lens

thickness results in two different projective centres C and c, for each of the two coordinate systems, respectively. However, angular relations between passing rays are preserved. Thus, only an extra translation and possibly rotation are involved, which may be assumed to be incorporated in (1–3). In effect, the projective relations involve directions rather than absolute points in two coordinate systems.

Condition Equations

For known exterior perspective centres, commonly denoted as *camera stations* with locations $\mathbf{X}_{C_i}$, and orientation matrices $\mathbf{M}_i$, the observed camera output signals $(x_{pi}, y_{pi})^T$ may serve to reconstruct the location $\mathbf{X}_P$ of a point P. A *photogrammetric triangulation procedure* basically consists of determining these external camera parameters by observing a number of control points with known locations. Subsequently, or simultaneously, all observed targets of interest may be reconstructed.

For these purposes, a set of equations relating observations, *a priori* information and unknown quantities must be obtained. One type of such *condition equations* follows by eliminating the proportionality factor k in (3), as achieved by dividing the first and second rows with the third, and by multiplying with the camera constant c:

$$\begin{pmatrix} x_p \\ y_p \end{pmatrix} = \frac{-c}{s_{p3}} \begin{pmatrix} s_{p1} \\ s_{p2} \end{pmatrix}, \tag{4}$$

with $\mathbf{s} = \mathbf{M}\,(\mathbf{X}_P - \mathbf{X}_C)$ observed direction in image coordinates.

These equations are called *collinearity equations* since P, p and C are on a common line.

In general, the condition equations may be concisely written as a set of implicit equations:

$$\mathbf{f}(\mathbf{x}; \mathbf{y}, \mathbf{b}) = 0, \tag{5}$$

with x unknown vector (e.g., target location in measurement); y observations (e.g., camera outputs); b system parameters (e.g., control point locations in calibration).

For a sufficient set of condition equations, and proper choice of control points or camera parameters, respectively, a unique solution may exist for the unknown quantities. In practice, the observations y and system parameters b will be subject to errors, and the solution will deviate from the true value. This effect may be reduced by employing a *redundant* set of observations, and the resulting set of equations (5) cannot be

solved exactly. An estimate $\hat{x}$ is subsequently sought which minimizes the equation error for vector f in some specific sense. The resulting error vector f ($\hat{x}$) is denoted as *residual* vector.

Least squares techniques are widely used for this purpose and will be discussed below; they have the additional advantage that *a priori* information, system parameters and actual observations may be treated in an essentially similar fashion, all being entered into condition equations with their respective reliabilities and interdependencies. For instance, one may account for known marker point distances on the human body, or for knowledge regarding the horizontality and length of the intercamera base of a stereometric camera [6]. A universal photogrammetric software procedure may be tailored to a particular adjustment problem by a proper choice of constraining interdependencies, weight factors, control points and camera parameters, all to be entered as data or as minor subroutines.

Systematic Errors in Photogrammetry (Second Order Theory)

The application of averaging techniques by fitting to a redundant set of condition equations is valid if the residuals are of a random nature. While some optimum may be found if certain systematic errors are not accounted for, subsequent analysis may indicate that additional improvements are feasible (fig. 2). Therefore, systematic deviations from the idealized, first order theory of the previous sections warrant a separate review, together with correction procedures. As indicated earlier, one may bring such procedures within the same scope, by expanding the analytical, photogrammetric model with a deterministic, parametric description of image distortion, and by viewing the remaining residuals as random.

Systematic errors may be categorized into errors of a global or local nature. Radial distortion caused by defocusing or lens errors has a typically global character, whereas flatness deviations in photographs during exposure [17] or nonhomogeneities in the resistance layer of lateral effect detectors [24] result in local image distortion. Since global errors often dominate deviations from ideal projective behaviour, local errors and their correction will not be discussed. Besides, the latter are more complicated and thus less interesting in real-time applications. An introduction to various procedures has been given by LEBERL [8].

The distinction between distortion and blurring of images adopted in the previous section looses its meaning for nonaddressable detectors [24]

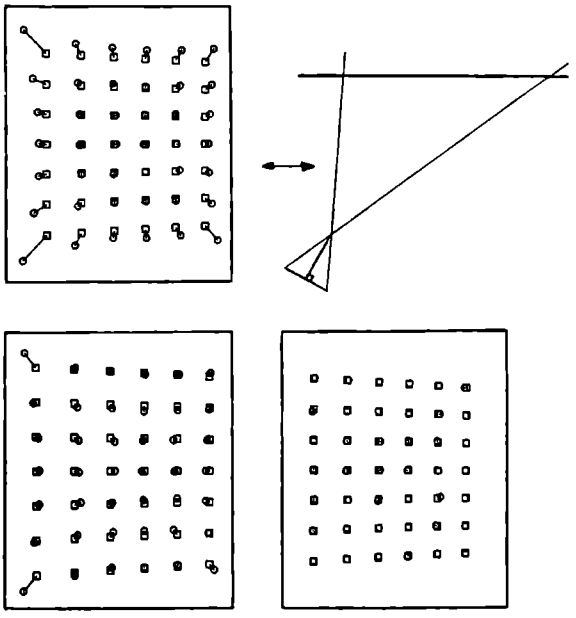


Fig. 2. Effect of systematic errors on image residuals. *Above left:* unadjusted; *below left:* external parameters adjusted; *below right:* internal and external parameters adjusted.

□ Image points from ideal, modeled camera.

O Observed image points with or without internal parameter adjustment.

which integrate the total incident light distribution. Thus, lateral detectors as employed in the SELSPOT system deliver the mean position of a projected light patch, irrespective of its shape or area, and output signals vary for both error types. As a consequence, the straight-line assumption of close-range light transmission which is fundamental to first order theory does not obtain: the influence of target-to-camera distance on distortion results in curved rather than straight loci for constant camera output signals. Again within a given range, such deviations may remain insignificant.

Focal length and distortion depend on light wavelength. By employing monochromatic light – either by filtering at the receiving end or by proper light generation – one may cancel this error source. For instance, the SELSPOT system employs narrow-band infrared LEDs together with infrared low-pass filters in front of the detectors; apart from the prevention

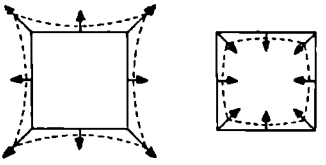


Fig. 3. Pincushion *(left)* and barrel *(right)* distortion.

of chromatic aberrations, the influence of background light on resolution [25] is thus reduced.

A dominating type of distortion results in radial image point displacement with respect to the principal point, as a consequence of lens errors. Two types are commonly distinguished: for outward displacement, the error is considered negative and called 'pincushion' distortion, whereas the opposite type is called 'barrel' distortion (fig. 3). This error may be well below 0.1% (full scale) if high-quality and thus costly optical systems are used [17]. Defocussing in nonaddressable detectors has a similar effect, and total radial distortion errors up to 3% full scale have been found in the SELSPOT system using large aperture closed-circuit TV-optics.

In a sense, one obtains a radius-dependent focal length; the term calibrated focal length denotes that particular value for c which minimizes the residual errors due to radial distortion in some specific sense. The effect of this calibration may be gleaned from figure 2.

Standard compensation procedures of higher complexity utilize interpolation tables or polynomial approximations in the radius r_p with respect to the principal point,

$$x_p' = r_p\,(1+\beta_1 \cdot r_p{}^T r_p + \beta_2\,(r_p{}^T r_p)^2 + \ldots), \tag{6}$$

with $r_p = \begin{pmatrix} x_p - x_o \\ y_p - y_o \end{pmatrix}$.

Improper centering of lens elements in a compound lens system may cause *tangential* distortion with image points displaced perpendicularly to the radius r_p. The effect is to curve image lines passing through the principal point [19].

Uniform film shrinkage during photograph development affects only the calibrated focal length, and cannot be viewed as a source of distortion. However, shrinkage may be different in various directions, and the amplification factors in x and y amplifiers of an optoelectronic system may dif-

fer; these effects entail distortion. By scaling the components of x_p' in (6), this error source may also be compensated,

$$x_p'' = \begin{bmatrix} 1-a & 0 \\ 0 & 1+a \end{bmatrix} \cdot x_p', \; |a| < 1. \tag{7}$$

All these errors have a global character, and their existence may often be concluded by visual inspection of a regular distribution of control point images. Other global deviations may be less evident but still amenable to global compensation procedures: the theory of linear prediction and filtering as developed by WIENER and KOLMOGOROV has in the last few years also been applied in photogrammetry [7].

Solution of Condition Equations

The estimation of an unknown vector x from an overdetermined system of implicit equations

$$f(x; y, b) = 0 \tag{5}$$

necessitates a quantitative definition of the notion 'optimal' or 'best' fit.

For deterministic problems, e.g. polynomial approximation of complicated mathematical functions, minimax criteria are often adopted; that is, a value $\hat{x}$ is chosen for which the largest component in f is as small as possible, for all admissable x. However, this criterium is not acceptable in the nondeterministic case since the minimax component may be large purely by chance. A better approach is to combine the various residuals into an integral criterium-function.

Standard statistical procedures include maximum likelihood methods and Bayes' decision theory [12, 18]; the former may be applied if the error distributions of y and b have a known parametric dependence on x, while the latter in addition requires an *a priori* known distribution function on x. Such information is usually lacking in photogrammetric problems, and the usual approach is to minimize a weighted norm of the residuals:

$$D_k = \left| \sum_{i=1}^{n} w_i \cdot | f_i(x; y, b) |^k \right|^{1/k}, \; k \geqslant 1 \tag{8}$$

(incidentally, the minimax criterium corresponds to the limiting case $k \to \infty$).

Least Squares Norm

For reasons of mathematical trackability, the quadratic or 'Euclidean' distance norm is usually employed:

$$D_2 = (f^T W f)^{1/2}, \tag{9}$$

where W is a weight matrix containing weight factors on its main diagonal for each observation equation. If the residuals are correlated, the off-diagonal elements may be chosen different from zero.

There are two reasons for the widespread use of this *least squares norm*. First, linear overdetermined equation systems

$$f(x) \equiv Ax + b = 0 \tag{10}$$

may be explicitly solved in terms of this criterium, as discussed in the Appendix. Secondly, least squares and maximum likelihood procedures render identical results if the error distributions are multivariate normal [12, 18], since all parametric information on this distribution is exploited in either method.

Explicit solvability generally does not obtain for nonlinear sets of equations. By linearizing the condition equations about an estimate x_i,

$$f + F_x \Delta x + F_y \Delta y + F_b \Delta b = 0, \tag{11}$$

with Δy and Δb unknown errors in y and b, respectively, $F_\cdot = \delta f / \delta_\cdot$ partial derivative of f to vector.

an optimal correction or *adjustment* $\Delta \hat{x}_i$ may be explicitly obtained on linear theory which minimizes the equation errors in (11), for some suitable weighting matrix W_i. Hopefully, the function value f_{i+1} will be smaller than the previous value f_i, in the sense

$$D_{2_1}^2{}_{+1} \equiv f_{i+1}^T W_{i+1} f_{i+1} < D_{2_1}^2 \equiv f_i^T W_i f_i, \tag{12}$$

with $x_{i+1} = x_i + \Delta x_i$.

By repeated linearization, solution of linear equations, and adjustment of a current estimate, an optimal value x may be found for which no further reduction in f is possible. Standard termination criteria for convergence include bounds on the decrement in D_2, and on the adjustment length $|\Delta x_i|$.

Convergence cannot be guaranteed for the general, nonlinear case: the criterium fuction may decrease even further outside some local environment of $\hat{x}_i$. For sufficiently strong photogrammetric models – that is, choice of control points and realistic mathematical description of the de-

terministic and noisy properties of the cameras – such local optima tend to be absent or at least not to hinder the iteration process Their presence may be verified to some extent by starting from different initial guesses $\hat{\mathbf{x}}_{i=0}$ if the final value of D_2 seems inordinately large.

Δy and Δb denote unknown errors in the observed y and *a priori* known b, respectively, with expected values 0 (other values would imply systematic errors which are assumed to be absent). If information regarding their reliabilities and interdependencies is absent, the weight matrices W_i are usually taken to be identity matrices, and *a least squares estimate* $\Delta\hat{\mathbf{x}}_i$ is obtained.

Usually, more information is available. Thus, in the case that f denotes image residuals, one may know when these are likely to be large. For instance, resolution in the SELSPOT system is ultimately limited by quantization noise in an AD-convertor. Practical limits are caused by the lateral detector's signal-to-noise ratio which is directly proportional to the intercepted light power [25]. This depends in turn on the angle of incidence and on light source distance, suggesting the introduction of certain weight factors in dynamic measurement procedures. During static operation as in calibration, a simpler method might be to average a sufficiently long series of observations.

In general, one will be able to postulate *(a priori)* or estimate *(a posteriori)* covariance matrices Y for Δy and B for Δb. In calibration, Y may be the identity matrix if image residuals on control points have equal standard deviations and no interdependency; if the control point locations are known with sufficient accuracy, there are no noisy system parameters b. During subsequent measurement, Y remains the same or depends on (only approximately known) target distance, while B becomes the covariance matrix which was the *a posteriori* result of the previous calibration procedure.

Based on linear theory (see Appendix), the optimal weighting matrix for (12) may be found to be

$$W_i = \left\{ \begin{bmatrix} F_y & F_b \end{bmatrix} \begin{bmatrix} Y & O \\ O & B \end{bmatrix} \begin{bmatrix} F_y & F_b \end{bmatrix}^T \right\}^{-1}, \tag{13}$$

if y and b errors are uncorrelated. Substitution in (12) yields a *minimum variance, weighted least squares estimate* for the adjustment step

$$\Delta\hat{\mathbf{x}}_i = -\operatorname{cov}(\Delta\hat{\mathbf{x}}_i) \quad F_{xi}{}^T W_i f_i, \tag{14}$$

with $\operatorname{cov}(\Delta\hat{\mathbf{x}}_i) = (F_{xi}{}^T W_i F_{xi})^{-1}$ correction covariance matrix

The residual vector $f(\hat{x})$ is usually quite small, and this renders linearization about the true and unknown value of x valid for the full range of possible errors in y and b. In that case, $\hat{x}$ is an *unbiased estimate* (i. e., the statistically expected value equals the true value), and cov $(\hat{x})$ equals the covariance matrix of the latest step $\lambda\hat{x}_i$ (14).

For crude initial guesses, the nonlinear character of the condition equations may cause failure of the convergence process. Thus, in resection a modelled image plane may originally bisect the control point distribution. If during adjustment the modelled image plane should pass through a control point, s_{p3} in (4) would pass through zero, and a criterium function expressed in image residuals would pass through a singular point. Such problems are unlikely to occur if one exercises some care in choosing initial guesses: a fairly accurate initial guess may often be obtained by visual inspection.

From (14) one may establish the final reconstruction accuracy in a consecutive series of calibration and measurement, by estimating the covariance matrix for an estimated target through (14), once the camera parameters and their covariance matrix have been determined.

Implicit and Explicit Estimators: Choice of Condition Functions

The common and theoretically most justified norm for optimization is to choose the condition functions f in terms of *image* residuals. Both calibration and measurement then require iterative, time-consuming procedures.

An alternative approach is to employ control point errors in *object* space as residuals, that is, the distance from a known control point location to an observed spatial direction is used. Such object space criteria have been used in instrumental photogrammetry since some instruments are based upon them [19]. In addition, target measurement may be carried out explicitly on this criterium, allowing shorter calculation times as required for real-time monitoring. This estimator may be derived as follows.

Assuming the camera parameters error-free, the observed directions S_j expressed in object space coordinates follow from (4):

$$S_j = M_j^T \begin{pmatrix} x_j \\ y_j \\ -c_j \end{pmatrix}, \quad j = 1, 2, \ldots, n; \; n \geqslant 2. \tag{15}$$

The distance vector D_j from any spatial point U at X_u to a line with direction S_j through camera station X_{Cj} follows as $D_j = P_j (X_u - X_{Cj})$, with $P_j = I^{3\times3} - S_j (S_j^T S_j)^{-1} S^T_j$ projection matrix onto a plane perpendicular to S_j.

P_j has the properties of idempotency and symmetry,

$$P_j^T P_j = P_j,\ P_j^T = P_j. \tag{16}$$

Defining a criterium function,

$$D_2^2(X_u) \triangleq \sum_{j=1}^{n} D_j^T D_j, \tag{17}$$

a spatial location $\hat{X}_u$ exists which minimizes D_2: the gradient to this estimated position should equal zero, which results in

$$\nabla_U D_2^2(X_u) = \sum_j P_j (X_u - X_{Cj}) = 0 \tag{18}$$

with a unique and explicit *least squares* solution

$$\hat{X}_u = (\sum_j P_j)^{-1} \sum_j P_j \ X_{Cj} \tag{19}$$

provided that the sum of the projection matrices is invertible. One may prove that this obtains if at least two directions S_j are not parallel.

For two cameras, relation (19) may be derived by geometrical considerations which results in a computationally simpler form. The S_j will generally cross at a finite distance, and the least squares estimate is located at the midpoint of the perpendicular connecting the two directions,

$$\hat{X}_u = \tfrac{1}{2}(X_{C1} + X_{C2}) + \{(L^T S_2)S_1 + (L^T S_1)S_2\} / |S_1 \times S_2|^2, \tag{20}$$

with $L = \tfrac{1}{2}(X_{C1} - X_{C2}) \times (S_1 \times S_2)$.

Residual error follows from the interconnecting perpendicular, which is in fact the projection of the intercamera base on the vector product of the two observed directions, with length:

$$|\Delta X_u| = (X_{C1} - X_{C2})^T (S_1 \times S_2)/|S_1 \times S_2|. \tag{21}$$

The algorithm defined by (15) and (20) has been programmed in Assembler language for use in a PDP-11/45 computer with floating point hardware. The calculation times ranged between 0.9 and 1.1 msec, for various combinations of observations and camera parameters. Since the SELSPOT system requires 3.1 msec to observe up to 30 LEDs, one may conclude that real-time reconstruction of target positions is fairly limited if the full system data rate is exploited. However, there are many examples in biomechancal research in which lower data rates or just a few targets are required.

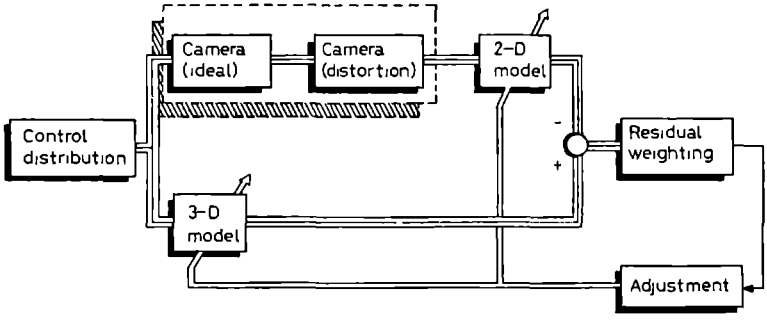


Fig. 4. Simultaneous estimation of internal and external camera parameters.

In the above algorithms, the camera parameters were assumed to be accurately known, and covariance matrices based upon these algorithms do not reflect the total system accuracy. For a complete error analysis *a posteriori,* the parameter errors should be taken into account, and this requires some insight in the control point distribution during calibration, as discussed in the next section.

In calibration, iterative procedures are usually unavoidable, and the condition equations will contain correction algorithms for image distortion and principal point location. Since these algorithms may contain polynomial approximations such as (6), to be used in future intersection operations, it is advisable to define condition equations in terms of differences between camera output signals after distortion correction and model outputs from ideal pinhole cameras, denoted as 2-D model and 3-D model in figure 4, respectively. The distortion compensation procedure is thus assessed in the form in which it will be used later; unlike the 3-D model, it cannot be easily inverted.

For rather accurate initial guesses, one iteration may suffice, and the distinction between implicit and explicit estimators becomes less meaningful. Thus, in dynamic movement monitoring, previously estimated positions may be used as or extrapolated to initial guesses for current positions. Yet, relations (15) and (20) would appear to be computationally simpler.

Because of the noisy character of observations on dynamic phenomena, a reconstructed 3-D trajectory will be far from smooth, and some form of data smoothing or filtering will be necessary in addition to distortion compensation. One approach is to filter the raw observation data

which may suffice if real-time requirements prevail. However, a theoretically sounder approach is to optimally combine new observations with previous trajectory estimates. The application of dynamic adjustment methods such as Kalman filters and related procedures is promising; the nonlinear character and complexity of the pertinent calculations are quite challenging, though, especially in a real-time context as will be discussed in Part II [26].

Geometrical Strength of Calibration and Measurement Models

In the following sections, the influence of control point distribution on camera parameter estimation, and of external camera parameters on reconstruction accuracy will be discussed.

Parameter Sensitivity to Control Point Distribution

The high correlation in resection between external camera parameters (including focal length) for a single control plane is illustrated in figure 5. Circles correspond to observed camera outputs, and squares to the optimally adjusted model outputs. The left picture shows the residual errors for the true camera parameters, the right picture those for optimally adjusted camera attitude and focal length, for faulty camera station. For a 3-D control point distribution, the observed image residuals would have shown a much greater discrepancy.

The correlation between camera attitude and station may be removed by a proper choice of control points [16]. For camera station calibration, the proper procedure is to adjoin a point A′ to a point A both yielding the same observation $a = (a_1, a_2)^T$. On first order theory, the camera station C is somewhere on the line passing through A and A′. By repeating this procedure on a control point B with corresponding B′ and $b = (b_1, b_2)^T$, $b \neq a$, space intersection may be explicitly carried out on the camera station, following the argument of the previous section. This approach necessitates position determination of arbitrarily chosen control points, which may be more complicated than space resection based upon a number of known control planes. A conclusion would seem to be that observations on different control planes should result in overlapping images. Yet, other correlations will persist: the high correlations between polynomial regression factors in (6) constitute an example. If internal and external camera parameters are simultaneously estimated, a

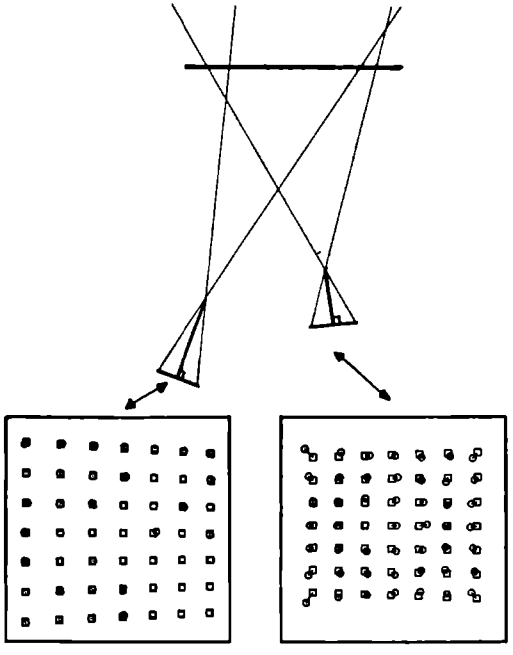


Fig. 5. Correlation between external camera parameters (including focal length) for a single control plane.

set of observations with image points which are well distributed over the image plane will show whether global correction procedures are sufficient. Distortion dependency on target distance may be checked by comparing residual errors for 3-D control point distributions with those obtained for a single control plane, after 'fixing' redundant parameters in the latter case.

The large amount of observations necessary for distortion correction 'in the field' may render simultaneous estimation of internal and external parameters rather costly or impractical, if active light sources are used. Thus, the LED light sources required for the SELSPOT system are quite expensive, and a cheaper *but* labourious alternative is to move a single LED between control locations in a control plane. One might therefore prefer to assess distortion parameters for a set of typical observation distances and lens aperture settings, to be used as standard correction procedures in both space resection and intersection: the number of control points required for space resection only is much lower. Thus, one might

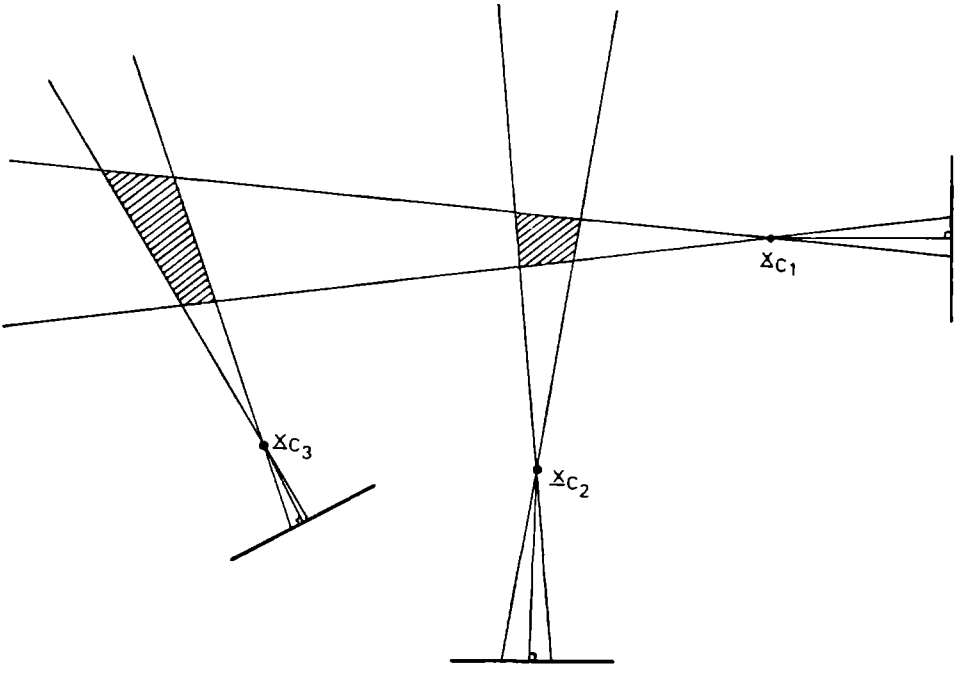


Fig. 6. Reconstruction accuracy dependence on parallactic angles and observation distance, for given angular observation accuracy.

display a simple 3-D control object consisting of a transparent cube with control points at its apexes.

Although system calibration is again performed by components rather than integrally in such an approach, the advantage of 'realistic' calibration circumstances is retained, since a calibration of internal parameters in the field may be repeated at will.

Accuracy in Space Intersection

An *a posteriori* accuracy analysis in space intersection by covariance assessment does not reveal much information on underlying error sources, unless a large set of parameter combinations is investigated. Some *a priori* considerations are presented in this section.

For given target distances, the observed directions should preferably intersect at approximately 90°, as illustrated in figure 6. For fixed intercamera base as is the case with stereometric cameras, and fixed target distance, it has been shown that accuracy is optimal for parallel optical axes

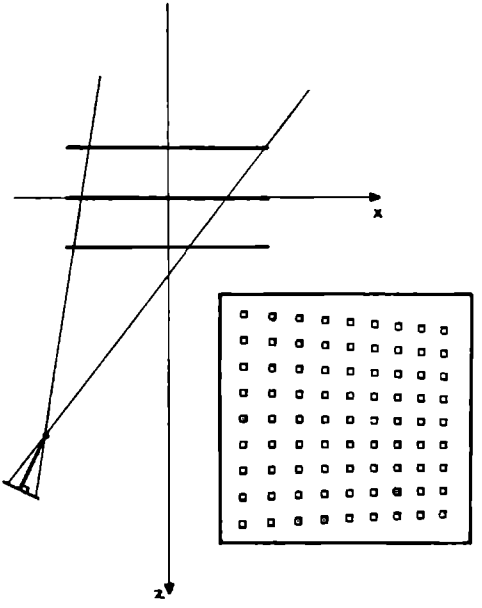


Fig. 7. Calibration of a SELSPOT camera: control plane and camera layout, and image residuals for the middle control plane.

[1]. However, this result is based on the assumption that angular resolution is determined by the transducer or subsequent equipment, and not by the amount of intercepted light as is the case with the SELSPOT system; if signal-to-noise ratio is determined by the incident light power, the optical axes should be oriented to the target area.

Accuracy in estimated positions will strongly depend on target distance, especially if light intensity is involved. For given light power at the target, the intercepted light power will vary inversely with the square of the target-to-camera distance, and angular resolution varies accordingly. In addition, figure 6 shows that spatial resolution varies inversely with observation distance, for constant angular accuracy, resulting in a cubic dependence of total positional reconstruction accuracy on observation distance.

Experimental Results on Simultaneous Calibration of Internal and External Camera Parameters

The results from a simultaneous adjustment of internal and external camera parameters of a SELSPOT camera are shown in figure 7. The

Table I

	x	σ_x	Unit	Cross-correlations, %												
				c	ω	φ	κ	C_x	C_y	C_z	β_1	β_2	β_3	x_0	y_0	a
c	54 570	0.188	mm	100												
External parameters																
ω	1.149	0.069	°	17	100											
φ	-22.784	0.070	°	-36	-17	100										
κ	0.123	0.035	°	17	79	-13	100									
C_x	-1.226	0.002	m	-62	-4	38	-3	100								
C_y	-0.025	0.001	m	-5	-23	7	-5	9	100							
C_z	2.345	0.005	m	62	4	-29	3	-91	-10	100						
Internal parameters																
β_1	-1.111×10^{-3}	7.6×10^{-5}	mm^{-2}	65	7	2	5	4	1	-5	100					
β_2	6.436×10^{-6}	7.2×10^{-7}	mm^{-4}	-58	-8	-2	-7	-5	-1	6	-97	100				
β_3	-1.371×10^{-8}	2.0×10^{-9}	mm^{-6}	51	10	4	8	7	1	-8	93	-98	100			
x_0	-0.948	0.069	mm	28	-20	-89	-16	-25	1	27	-6	6	-8	100		
y_0	0.026	0.068	mm	14	96	-16	80	-4	3	4	7	-9	10	-20	100	
a	0.007	0.001	-	-30	-69	70	-54	23	19	-21	-4	5	-5	-43	-66	100

camera was equipped with a Canon TV-16 fixed focus lens $f = 55/0.95$, and three control planes with totally 209 control points were employed as depicted. The distortion calibration procedure employed the relations (5) and (6), with three polynomial terms in r_p. Principal point x_o, differential scaling a and camera constant c were simultaneously estimated with camera attitude and station. The residual image error $\hat{o}$ (see Appendix) was 44 μm per axis for 24 mm full-scale range, which corresponds to 1.8 least significant bit for the 10 bits AD-convertor contained in the central camera processing unit. Estimated standard deviations and correlations are presented in table I.

For the estimated values of camera station and attitude, the internal parameters and focal length were adjusted per calibration plane. Since the residual standard deviation remained approximately the same when these parameters were subsequently used for the other control planes, one may conclude that the straight-line assumption on effective light transmission was valid for the given control volume. The high correlation between distortion parameters and parameter sensitivity on observations render constancy of estimated parameters less suitable as a criterion.

Suboptimal quality of the – preliminary – camera transducer is suspected to be a major error source. Present work is carried out on a system with detectors of a higher quality, and will be reported in part II of this paper [26], together with results on total system accuracy in static and dynamic intersection.

Appendix

Linear Least Squares

As clarified in the main body of this paper, calibration and measurement may be reduced to a repeated process of solving an overdetermined set of equations in a least squares sense:

$$F_x \Delta x = -f \tag{A1}$$

with

$f = f' + v$ current equation error;

$f' = f(\hat{x}_i) - f(x)$ discrepancy vector due to initial approximation $\hat{x}_i$ of the unknown parameters x;

$v = F_y \Delta y + F_b \Delta b$ error vector due to finite accuracy in observations y and *a priori* known parameters b.

As has been proven in many texts on statistical inference [12, 18], an optimal adjustment $\Delta\hat{x}$ in the sense of unbiasedness and minimum variance may be

obtained if the error vector v has a known expected value $\bar{v}$ (assumed to be zero since a nonzero value would correspond to disregarded systematic errors) and covariance matrix V. If V is unknown and some other covariance matrix is assumed, unbiasedness is retained, but the *a posteriori* variance of the estimate $\Delta\hat{x}$ will increase.

Premultiplying (A1) with a square root U of V^{-1}, i.e.:

$$U^TU = V^{-1} \tag{A2}$$

a standard least squares problem is obtained,

$$(UF_x)\,\Delta x = -Uf, \tag{A3}$$

where the error vector $v' = Uv$ has unit covariance $V' = I$, that is, each new condition equation is equally reliable and uncorrelated with other equations.

The least squares solution $\Delta\hat{x}$ follows analytically by forming the *normal equations:*

$$(UF_x)^T(UF_x)\,\Delta x = (UF_x)^T \cdot -Uf \tag{A4}$$

which by relation (A2) may be compressed into

$$(F_x^TV^{-1}F_x)\Delta x = -F_x^TV^{-1}f. \tag{A5}$$

If these equations are linearly independent – this depends on the control distribution or camera distribution – the bracketed term may be inverted, resulting in:

$$\Delta\hat{x} = -\mathrm{cov}(\Delta\hat{x})\,F_x^TV^{-1}f \tag{A6}$$

with covariance matrix

$$\mathrm{cov}(\Delta\hat{x}) = (F_x^TV^{-1}F_x)^{-1}$$

and weighted residual vector estimate

$$\hat{v}' = Uf + UF_x\Delta\hat{x}.$$

If the above matrix inversions are not possible, e.g. because of insufficiently strong control distribution, one may resort to *generalized inverses* [13]. In essence, these remove adjustments on parameters which are superfluous because of perfect correlations with other parameters. The resulting estimate $\hat{x}$ may be wrong for observations on points not contained in the original control subspace (e.g. a plane), but $\hat{x}$ may be quite acceptable if observations after calibration are confined to this subspace.

In many instances, the absolute *a priori* covariance matrix V is not known, only relative values being available,

$$V = \sigma^2\Sigma, \text{ with } \Sigma \text{ known.} \tag{A7}$$

Thus, in resection and internal calibration, the observations may be equally reliable and uncorrelated – assuming little or no image distortion – which results in $\Sigma = I$. Using Σ rather than V in the above relations, an unbiased estimate $\hat{\sigma}^2$ for σ^2 follows from the weighted sum of squared residuals

$$\hat{\sigma}^2 = f^T \Sigma^{-1} f/(m-n),$$

where m is the number of condition equations, and n the number of parameters to be estimated.

When convergence in repeating (A6) has been attained ($\Delta\hat{x} \to O$), the statistics $\mathrm{cov}(\Delta\hat{x})$ and $\hat{\sigma}^2$ do not only apply to the adjustment $\Delta\hat{x}$, but also to the final estimate $\hat{x}$ (cf. table I).

Although (A6) follows explicitly by inverting a matrix product, rounding errors in numerical work may render the solution rather unreliable, especially if residuals vary over a wide range or if the equations are highly correlated. This may deteriorate the convergence rate or even result in divergence. Numerically more stable procedures utilize *orthogonalization* methods [13, 21], that is, the left and right hand sides of (A1) are premultiplied with an orthogonal matrix Q (i.e., $QTQ = I$), so that the resulting equations

$$(QF_x)\Delta x = -Qf \tag{A9}$$

are easier to solve. The residual sum of squares is not affected by this operation, since

$$(Qf)^T(Qf) \equiv f^T(Q^TQ)f \equiv f^Tf. \tag{A10}$$

For a suitably chosen matrix Q, the product QF_x may be rendered of upper triangular form R.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \dots \\ 0 & r_{22} & r_{23} & \dots \\ 0 & 0 & r_{33} & \dots \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{bmatrix} = \begin{bmatrix} R^* \\ O \end{bmatrix} \tag{A12}$$

Matrix inversion and solving in a least squares sense are reduced to a numerically stable process of back substitution, e.g.,

$$\begin{aligned} r_{11}x_1 + r_{12}x_2 + r_{13}x_3 &= y_1 \\ r_{22}x_2 + r_{23}x_3 &= y_2 \\ r_{33}x_3 &= y_3. \end{aligned} \tag{A13}$$

The equations are solved by starting with x_3, substituting, solving for x_2, substituting, etc.

The orthogonal matrix Q may be obtained by certain operations on the original matrix F_x. Standard orthogonalization procedures such as (modified) Gram-Schmidt and Householder transformations [13, 21] have been used in photogrammetric least squares adjustment, with good results [27]. They exhibit a disadvantage insofar they operate columnwise on the matrix F_x. Since the condition equations are usually highly overdetermined ($m \gg n$) in order to reduce the influence of observation and parameter errors, one may run into memory storage problems. One remedy is to rely on backing storage such as magnetic drum or disk with a consequent increase in programme run time. Another method which has been employed in this investigation is to orthogonalize the condition matrix F_x in a rowwise fashion. GIVENS [4]

and GENTLEMAN [3] have described procedures for diagonal weighting matrices requiring storage room in the order of $^1/_2n^2$ rather than m·n as needed for column-wise operations.

Data may be fetched or generated row by row, each observation equation being required only once per iteration step (11). Especially in a nonlinear, iterative context, rowwise orthogonalization might be preferable since the partial derivatives in F_x must be calculated at each step, with corresponding elements in a single row.

References

1 ABDEL-AZIZ, Y. I.: Expected accuracy of convergent photos. Photogramm. Engng *40:* 1341–1346 (1974).

2 BUDINGER, T. F. and GULLBERG, G. T.: 3-D Reconstruction in nuclear medicine emission imaging. IEEE Trans. Nucl. Sci. *21/3:* 2–20 (1974).

3 GENTLEMAN, W. M.: Least squares computations by Givens transformations without square roots. J. Inst. Maths Applics *12:* 329–236 (1973).

4 GIVENS, W.: Computation of plane unitary rotations transforming a general matrix to triangular form. *J. SIAM 6: 26–50 (1958).*

5 GORDON, R. and HERMAN, G. T.: 3-D reconstruction from projections: a review of algorithms. Int. Rev. Cytol. *38:* 111–151 (1974).

6 KRATKY, V.: Digital modelling of limbs in orthopaedics. Photogramm. Engng Rem. Sens. *41:* 741–752 (1975).

7 KRAUS, K. and MICHAEL, E. M.: Linear least squares interpolation. Photogramm. Engng *38:* 1016–1029 (1972).

8 LEBERL, F.: Photogrammetric interpolation. Photogramm. Engng Rem. Sens. *41:* 603–612 (1975).

9 LINDHOLM, L.-E. and ÖBERG, K. E. T.: An opto-electronic instrument for remote on-line movement monitoring; in NEUKOMM Biotelemetry II, 100–102 (Karger, Basel 1974).

10 LIPPERT, F. G.: The feasibility of photogrammetry as a clinical research tool. J. Biomech. *6:* 459–473 (1973).

11 MERCHANT, D. C.: Calibration of the air photo system. Photogramm. Engng *40:* 605–617 (1974).

12 MOOD, A. M. and GRAYBILL, F. A.: Introduction to the theory of statistics (McGraw-Hill, New York 1950, 1963).

13 PETERS, G. and WILKINSON, G. H.: The least squares problem and pseudoinverses. Comput. J. *13:* 309–316 (1970).

14 Photodetector picks up LED emissions to measure positions of moving objects. Electronics *48/4: 6E–8E* (1975).

15 PLAGENHOEFF, S.: Patterns of human motion: a cinematographic analysis (Prentice-Hall, Englewood Cliffs 1971).

16 RAMPAL, K. K.: Optimum ground control for camera calibration. Photogramm. Engng Rem. Sens. *41:* 113–118 (1975).

17 SCHERZ, J. P.: Errors in photogrammetry. Photogramm. Engng Rem. Sens. *41:* 493–500 (1975).

18 SILVEY, S. D.: Statistical inference. Penguin Library of University Mathematics series, Harmondsworth (1970).
19 THOMPSON, M. M.: Manual of photogrammetry; 3rd ed. (Am. Soc. Photogrammetry, Falls Church 1966).
20 WHITE, A. A., III, *et al.*: A system for defining position and motion of the human body parts. Med. Biol. Engng *13:* 261–265 (1975).
21 WILKINSON, J. H. and REINSCH, C.: Linear algebra. Handbook for automatic computation, vol. 2 (Springer, Berlin 1971).
22 WINTER, D. A., *et al.*: Tv-computer analysis of kinematics of human gait. Comp. biomed. Res. *5:* 498–504 (1972).
23 WOLF, P. R.: Elements of photogrammetry (McGraw-Hill, New York 1974).
24 WOLTRING, H. J.: New possibilities for human motion studies by real-time light spot position measurement. Biotelemetry *1/3:* 132–146 (1974).
25 WOLTRING, H. J.: Single- and dual-axis lateral photodetectors of rectangular shape. IEEE Trans. Electron. Dev. *22:* 581–590, 1101 (1975).
26 WOLTRING, H. J.: Calibration and measurement in 3-D monitoring of human motion by optoelectronic means. II. Experimental results and discussion. Biotelemetry *2* (1975), in print.
27 YASSA, G.: Orthogonal transformations in triangulation adjustment. Photogramm. Engng *40:* 961–966 (1974).

H. J. WOLTRING, Laboratory of Psychology, University of Nijmegen, Erasmuslaan 16, *NL–6804 Nijmegen* (The Netherlands)

Biotelemetry 3(1977)

It is the depth at which we live
and not at all the surface extension that imports.

R. W. Emerson, Society and Solicitude.

2.1.2

Calibration and Measurement in 3-Dimensional Monitoring of Human Motion by Optoelectronic Means

II. Experimental Results and Discussion[1]

H. J. WOLTRING

Laboratory of Psychology, University of Nijmegen, Nijmegen

Key words. Motion Study · Biomechanics · Displacement Sensing · Transducers · Software · Measuring Equipment · Photogrammetry · Kalman Filtering · Light Telemetry

Abstract. In the first part of this paper, the theoretical aspects of three-dimensional (3-D) point position reconstruction from 2-D observations through central projective imagery have been discussed. This *photogrammetric* theory has been applied to a 3-D version of a SELSPOT optoelectronic camera system for real-time direction monitoring of up to 30 IR LEDs. Some results on camera parameter estimation (e.g. position, attitude, radial distortion) and on static LED position reconstruction are presented, based on calibration with recourse to a regular, 3-D distribution of spatial *control points* whose locations are fully known ('absolute control'); in addition, some comparative data on 'temporal' and 'geometrical' reconstruction noise are given.

Since shadowing effects and body marker rotations may entail random interrupts of the incoming data stream, filtering procedures which process the raw camera signals may have to bypass unreliable observations; the *extended Kalman filter* and its suboptimal counterparts are quite appropriate for this purpose. In addition, time-domain methods such as Kalman filtering allow simultaneous estimation of position and its derivatives which are frequently required in biomechanics. An example of 3-D dynamic reconstruction of a human gait pattern is presented.

The calibration results suggested that additional improvements are feasible by employing more complicated distortion-compensation models; the application of general, polynomial approximation functions has appeared to be promising, as became apparent by assessing long-term system stability at high accuracy levels.

Finally, some alternative concepts for calibration and measurement are discussed, with emphasis on flexibility in calibration, and on calculation speed in measurement of human motion.

[1] Parts of this paper were presented under the same title at the 3rd International Symposium on Biotelemetry, Pacific Grove, Calif. 1976.

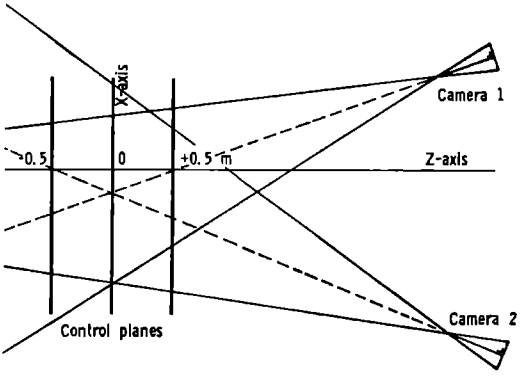


Fig. 1. Control and camera layout in calibration and static measurement.

Further Results on Camera Calibration

The calibration procedure outlined in Part I [49] has been applied to an improved version of the SELSPOT system [24], for a camera configuration as shown in figure 1. As before, the distance between the control planes was 0.5 m, and a rectangular area for each control plane was chosen in such a way as to maximize the image area covered for each observed *frame*. This procedure was repeated after interchanging the camera connections to the central signal-processing unit in order to verify whether image distortion was differentially influenced by this unit due to e.g. sample-and-hold limitations, different gain factors in preamplifiers prior to signal normalization [47, 48], etc. The resulting estimates and their *a posteriori* standard deviations are shown in table I, based on equal *a priori* weights for all condition equations. It should be recalled from Part I that most parameters are significantly correlated with each other.

The following conclusions have been drawn from these results.

(1): the distortion model for radial distortion and differential amplification as discussed in Part I yields a significantly better fit for camera No. 1. The residual standard deviations $\hat{\sigma}$ are better than those obtained prior to a recent system improvement with new lateral photodetectors: 1.6 and 2.4 LSB instead of 1.8 and 4.6 LSB, respectively. By covariance analysis between image residuals, e.g. as a function of the distance between image points, one may ascertain whether any systematic effects have not been sufficiently accounted for by the distortion model.

Table I. Estimated camera parameters and standard deviations

Symbol	Camera No. 1		Camera No. 2		Units
	channel A	channel B	channel B	channel A	
φ_C	0.47 (0.08)	0.47 (0.08)	−0.35 (0.11)	−0.18 (0.11)	degrees
	−22.13 (0.06)	−22.30 (0.06)	20.24 (0.08)	20.36 (0.08)	degrees
	−0.43 (0.04)	−0.42 (0.04)	−0.53 (0.05)	−0.58 (0.05)	degrees
X_C	−1,337 (3)	−1,317 (3)	799 (4)	813 (4)	mm
	−13 (1)	−15 (1)	−42 (2)	−46 (2)	mm
	2,839 (7)	2,778 (6)	2,705 (10)	2,732 (9)	mm
c	47.58 (0.16)	45.07 (0.15)	46.25 (0.22)	48.21 (0.23)	mm
x_o	0.34 (0.05)	0.34 (0.05)	−0.68 (0.07)	−0.67 (0.07)	mm
	−0.23 (0.06)	−0.29 (0.06)	−1.44 (0.08)	−1.38 (0.09)	mm
β	−344 (90)	−475 (92)	−696 (98)	−667 (92)	10^{-6} mm^{-2}
	−220 (110)	−70 (113)	158 (96)	143 (86)	10^{-8} mm^{-4}
	143 (37)	101 (41)	3 (27)	2 (22)	10^{-10} mm^{-6}
a	34 (1)	40 (1)	21 (1)	15 (1)	10^{-3}
$\hat{\sigma}$	38.3	37.1	57.8	59.0	μm
	1.6	1.6	2.4	2.5	LSB

Canon TV-16 fixed focus lenses 50 mm/0.95–22; aperture setting, 1.4; focal setting, 3 m (IR mark). SELSPOT data averaged over 1 sec (330 observations per point); LED No. 1, approximately 200 control point positions.

(2): the central signal-processing unit apparently exhibits significant differential distortion not modelled by differential scaling, since the changes in various correlated parameters are larger than one would expect on the basis of the *a posteriori* standard deviations. However, the parametric model for ideal cameras and image distortion was sufficiently general to account for these changes, since the residual standard deviations remained approximately the same for each camera.

Estimated internal parameters have been used to calculate radial distortion as a function of image radius. Effective distortion (the result of both lens errors and defocussing) has been found to be of the order of 3% full-scale for image radii ranging up to 80% full-scale when referred to an extrapolated central image which is assumed distortion-free. Minor variations may occur, since radial distortion under close-range circumstances varies with target distance, lens-distance setting and lens aperture. For the con-

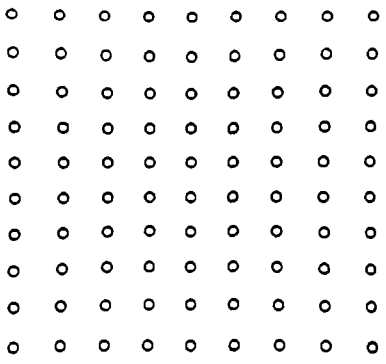


Fig. 2. Typical image distortion for Canon TV-16 fixed-focus lens 50 mm/0.95–22 commonly used in wide-range SELSPOT measurements. Aperture at 0.95; distance at 3 m (IR mark).

figuration of figure 1, and for the residual errors presented in table I, variations due to target distance were found to be negligibly small, as was the case with the preliminary observations in Part I.

A typical distorted image as obtained from a uniform, rectangular control distribution is shown in figure 2; the camera was oriented with its image axes approximately in parallel with those of the calibration plane. The Canon lens clearly exhibited strong pincushion distortion, which may be slightly compensated by choosing lens-distance settings which are actually too large: the blurred image will shift toward the optical camera axis since the distance between image plane and the lens is reduced in this manner. In addition, any local irregularities in the optoelectronic position-conversion process are smoothed; it would appear, however, that such local errors are negligible in currently available SELSPOT detectors.

That radial distortion is mainly caused by the Canon lens and not by the lateral photodetectors in the SELSPOT cameras may be concluded from the following. (1) Decreasing lens aperture has a beneficial effect on image quality. (2) Calibration curves obtained from the manufacturer of the lateral photodetectors in the SELSPOT cameras show virtually complete position rectilinearity, as assessed by moving an LED-equipped optical fiber over the position-sensitive detector area. This is a marked improvement with respect to the theoretical [48] and experimental [47] position characteristics of the tetra-lateral photodiode with rectilinear lateral contacts at one junction side, which was employed in our own experimental system and in the SELSPOT prototype [24].

Consequently, the use of more costly, photogrammetric lenses for measurement purposes may reduce image distortion to a level which is compatible with camera noise as will be discussed in a following section, although the occurrence of defocussing due to distance variations in close-range observations may remain a significant distortion source.

For the given polynomial distortion model with three parameters as discussed in Part I, residual image error increases with increasing image coverage. As an example, parameter estimation for $80 \times 80\%$ image coverage yielded $\hat{\sigma} = 2.3$ LSB, whereas 1.5 LSB was found for $65 \times 65\%$ coverage, with the camera at 2.2 m distance from the control plane, and for an equal number of observations; lens-distance setting was at infinity (IR mark), and the aperture was set at 0.95. In addition, image errors become increasingly dependent on target distance if larger distance variations occur. Thus, in one experiment it was attempted to employ absolute control for two remote control planes (3.5 and 4.5 m from the camera, respectively) which only partially covered the image, in combination with *partial control* for a nearer control plane allowing $90 \times 90\%$ image coverage, at approximately 2.1 m from the camera. Position and attitude were unknown variables which were adjusted with the camera parameters in the calibration procedure. Residual image standard deviation for all control planes was 3.3 LSB, for the two absolute control planes 1.1 LSB, and for the near control plane (fixing redundant external parameters) 2.2 LSB.

Static Reconstruction of Observed Targets

The above camery-parameter accuracy analysis is only an intermediate step, though a crucial one, toward a final analysis of error propagation in 3-D position and trajectory estimation. Some tests have been run on steady targets by employing the implicit and explicit estimators discussed in Part I, in order to assess steady-state system performance capabilities. For the explicit or *one-shot* estimator, the camera-parameters were considered error-free. As discussed in the following section, the results in dynamic reconstruction are less accurate due to camera noise.

Since the observed LED targets were displayed through the control plane originally used in calibration, absolute error evaluation in one control plane position was possible. As an example, the data in table II refer to positions in the rear control plane ($Z = 0.5$ m) of the configuration depicted in figure 1. This constitutes a worst case example, since the target-to-camera distances

Table II. Two typical examples of reconstruction accuracy

	True position, mm	Explicit estimate error, mm	Implicit estimate error, SD, and corr. matr.
X	0	-1.95	-1.94 (2.60) mm 100%
Y	0	-3.79	-4.13 (2.64) mm -1% 100%
Z	-500	2.01	2.03 (8.08) mm -4% -2% 100%
Residual	-	4.09	40 μm
X	-600	-2.45	-2.45 (2.55) mm 100%
Y	0	-5.30	-5.30 (2.63) mm -1% 100%
Z	-500	-3.08	-3.08 (7.84) mm 15% -2% 100%
Residual	-	4.13	40 μm

Camera No. 1 to channel B; camera No. 2 to channel A. Parameters as in table I, 64 averaged LED observations per position. Explicit residuals: length of interconnecting perpendicular; implicit residuals: residual image error.

were largest, and since the parallactic angles were smallest for this control plane.

The Y-components exhibit a systematically larger absolute error than one would expect on the basis of the *a posteriori* reconstruction standard deviation; since the camera y-axes were approximately in parallel to the object space Y-axis, this finding suggested that some submerged error in the y-channel of a least one camera was not sufficiently accounted for in the distortion-compensation model. This tallies with the larger residual errors for camera No. 2 as shown in table I. (By using a borrowed camera of more recent vintage instead of camera No. 2, residual image variances became approximately equal to each other.)

However, relative accuracy within this control plane, after compensating such observations with the mean error value, was quite consistent with the estimated *a posteriori* standard deviations.

In all these static calibration and measurement procedures, camera output signals are assumed invariant with the intercepted light intensity, between the SELSPOT overload and out-of-range light intensity levels. For dynamic measurements as discussed in the following sections, this is certainly not true. For the static case, where system noise is assumed to become negligible by averaging a sufficient number of raw camera observations prior to reconstruction of the target position, an other error source persists: offset

errors in preamplifiers and sample-and-hold circuits prior to signal normalization [47, 48] for light intensity compensation will have an adverse effect, and this must be countered by hardware readjustment once in a while. In addition, system performance becomes rather erratic when the intercepted light intensity falls to the out-of-range level: camera noise in the light-intensity compensation procedure reduces spatial resolution substantially. This will be further discussed in a later section.

Dynamic Reconstruction

Based upon the physically oriented camera and image-distortion models above, some experiments on dynamic reconstruction from single, raw camera observations have been carried out. Calibration for the configuration as shown in figure 3 took place along similar lines as above. A different configuration was chosen in order to have a sufficiently large field of view for the gait observations discussed in the next section. Residual image standard deviations were 1 LSB for camera No. 1 (248 control points), and 1.6 LSB for camera No. 2 (272 control points).

After calibration, the mean position and standard deviations of stationary LEDs were calculated, for 192 observations, using the explicit, one-shot estimator as above, and compared with theoretically expected values if quantization noise in the 10-bits AD-convertor of the SELSPOT central signal-processing unit were the only error source. Under the assumption of a uniform noise distribution with a range of 1 LSB per axis, the quantization noise standard deviation per axis may be calculated as $\sqrt{1/12}$, or approximately 0.3 LSB; this value was entered into the *a priori* observation covariance matrix of the iterated, implicit estimator discussed in Part I.

A 1 m long bar with 10 LEDs mounted on it was vertically placed at the positions F, M, R and W indicated in figure 3; the observed and calculated standard deviations in the LED positions are shown in table III. Geometrical distortion is of no importance in these observations since the targets are stationary, with interest focused on camera noise.

As apparent from the data in table III, the experimental standard deviations tend to be larger than the theoretically expected values; in addition, they increase more rapidly with distance to the cameras then is consistent with the assumption of constant camera noise. The latter may be explained with reference to the decreasing intercepted light intensity, as discussed in Part I.

Table III. Observed and expected, relative accuracy in dynamic reconstruction (SELSPOT overload at Z ≃ + 0.5 m)

Position	Depth Z m	Standard deviation, mm					
		σ_X		σ_Y		σ_Z	
		obs.	exp.	obs.	exp.	obs.	exp.
Front	0.0	0.6–0.8	0.6	1.0–1.3	0.5	1.9–2.4	1.4
Mid	–0.5	0.7–1.0	0.6	1.5–2.3	0.6	1.6–2.6	1.6
Rear	–1.0	0.6–1.3	0.7	1.7–3.0	0.7	1.8–3.2	2.0
Wall	–1.6	1.0–2.1	0.8	2.7–5.0	0.8	2.5–5.4	2.4

LED markers used: SELSPOT power LEDs.

The noise in the Y-direction is substantially greater than in the X-direction. This appeared again due to the y-channel of camera No. 2, as verified by direct camera-output signal observation.

The utility of any model for geometric image distortion calibration is limited by these noise effects. For the present model of radial distortion and differential scaling, the effects of 'temporal' and 'geometrical' ('spatial') noise on relative reconstruction accuracy have been compared, by calculating mean values and standard deviations of the distances between the LEDs on the bar. For position M in figure 3, and vertically positioned bar, the distance standard deviations varied between 0.5 and 1.5 mm (the additional noise in the y-channel of camera No. 2 was rather low-frequent, apparently, since these values are smaller than one would expect from the individual LED data in table III). When the bar was moved within the calibration area, with the LEDs oriented toward the cameras in order to maximize the intercepted light intensity, the standard deviations were found to increase to 2–5 mm, which is substantially more than one would expect on the basis of varying observation distance.

In this experiment, the bar was moved perpendicularly to its own orientation, so as to minimize influence of the time lag between SELSPOT observations on the various LEDs. For instance, in the case of movement at 1 m/sec in the direction of the bar, a time lag of 1 msec as between LED No. 1 and LED No. 8 would result in an apparent distance change of 1 mm.

These results suggest two different things: first, raw observations must be filtered or smoothed, especially if they are to serve for velocity and ac-

celeration estimation; and second, additional image distortion compensation seems useful. These topics are discussed in the following two sections.

Further Processing of Noisy Position Observations: an Application of Kalman Filtering

In many biomechanical applications, some further processing of raw direction or position observations is required. Usually, one is also interested in other *kinematic* functions such as velocity and acceleration; sometimes, even the third derivative ('jerk') is sought. By combining this kinematic information with certain anthropometric parameters [4, 7, 10, 15], *kinetic* motion analysis in terms of forces and moments at limb joints [35], and of energy flow across joints [46] becomes feasible.

The estimation of smooth trajectories and their time derivatives from noisy direction or distance observations is a classical problem: GAUSS [12] applied least-squares curve-fitting techniques to planetary motion almost a century ago, and his work has proven germinative to all subsequent filtering and smoothing methods. A major development such as Wiener's work in the frequency-domain was aimed at related problems: prediction of target locations in fire control. Similarly, the subsequent work of KALMAN and BUCY in the time-domain has elicited a host of applications including navigation, missile guidance, orbit determination, and fire control.

The notion *filtering* formally refers to estimating current phenomena from current and past observations, whereas *smoothing* denotes a 'post mortem' analysis including future observations with respect to the time-instant of interest. Smoothing may be subdivided into *fixed-point smoothing*, where a current estimate is improved with further future observations, *fixed-lag smoothing*, where an estimate is based on past and current information, and on a fixed amount of future observations, and *fixed-interval smoothing*, where all data within a given measurement interval are used to obtain estimates for arbitrary time-instants within the interval [18, 28]. *Prediction*, finally, refers to estimation of future phenomena from current and past observations. Traditionally, the term filtering includes both prediction and smoothing as in the term 'digital filtering'; the more strict meaning defined above will be adhered to, unless the opposite appears to be contextually the case, as in the following paragraph.

Common filtering procedures, some of which have been used in biomechanics include polynomial curve fitting followed by finite differencing

[29] or analytical differentiation [35] for derivative assessment, Fourier analysis with smoothing and differentiation occurring in the frequency domain [42, 47], and recursive or nonrecursive digital filtering in the time-domain [27, 33, 34, 36] where the filter is usually synthesized via frequency-domain considerations. An alternative approach is based on inverse dynamic programming [9], where assumed accelerations are iteratively adjusted over the whole measurement interval so as to minimize a quadratic cost function of the difference between the observed and integrated position data. PEZZACK *et al.* [34] recently presented a comparative study of various methods, comparing differentiated position data with actual accelerometric observations.

Especially acceleration estimation from noisy position observations is a very unreliable process, unless the position data are filtered to a narrow bandwidth. Some methods are quite acceptable for position smoothing but unreliable for differentiation, such as polynomial curve fitting: there is no guarantee that finite differences or analytical derivatives will properly reflect true velocities and accelerations if these signal functions are not included in the cost function to be minimized. PEZZACK *et al.* [34] have amply demonstrated this effect.

Fourier analysis and digital filtering where the filter is specified in terms of frequency responses are essentially stationary operations which require a bandwidth-determined settling interval. The use of such approaches implies conceptual modelling of a motion as a series of sinusoidal components. For nonperiodic motion this is an equally arbitrary choice as curve-fitting with polynomial functions, although more reliable from a numerical point of view. However, many types of extended motion may be considered as stationary, with major low-frequency components, whereas measurement equipment may exhibit dominant high-frequency noise. Thus, sinusoidal analysis with ordinary low-pass filtering will often give good results.

At the same time, however, the occurrence of random observation interrupts due to, e.g. shadowing effects or optical misalignment of LED body markers in the case of the SELSPOT system may necessitate frequent filter reinitializations. Moreover, sinusoidal filters are typically of the single-input/single-output type; vector-valued input signals such as x-y signals from various cameras and output signals such as simultaneously estimated positions and their derivatives in 3-D are hardly accounted for.

Recent advances in time-domain filtering as initiated especially by KALMAN's and BUCY's work in the early 1960s avoid all these drawbacks: thus, one may simultaneously estimate positions and higher derivatives from

multiple-sensor signals, and bypass unreliable sensor data [19, 39, 40]. In addition, these filtering procedures may be naturally extended to time-varying, nonlinear processes and measurements [16, 25], and to prediction and smoothing [18, 28, 32]. In essence, the time-domain description allows direct modelling of the observed process in terms of all variables of interest: a *state vector* (see Appendix) containing for instance position and higher derivatives or kinetic elements is continuously adjusted or *updated* by new, incoming information. In the case of a state vector with kinematic elements, one has a local polynomial approximation of the movement process with time as the independent variable [32]. If the observational data stream is interrupted, this model may be left 'freewheeling' until the information stream is resumed: the filter acts as a predictor.

In a similar vein as for the image distortion model discussed before, errors in the (limited) process model are considered as noise with known covariance matrix. Thus, a covariance matrix of the state vector may be estimated together with the state elements themselves, and new observations with known or assumed covariance matrix may be used to adjust the state vector in a weighted least-squares or *minimum variance* sense.

It is worthwhile to note that time-domain methods may be naturally generalized to accommodate identification in multitarget situations when addressable detectors [47] are used [3, 41, 43]. Also, systematic discrepancies between modelled and observed data may be used to adjust the *process model* rather than the state vector; this approach is known as *adaptive filtering* [2, 16, 23], and is a typical example of *joint state and parameter estimation* [11].

It is beyond the scope of this paper to present an acceptable review of Kalman filtering and related procedures; good introductions to various aspects of time-domain methods are available [11, 14, 21, 32], and more advanced coverage from a statistical or probabalistic point of view may be found [11, 16, 22, 25]. A summary is given in the Appendix, together with data on actual filter parameters as used for the experiment reported below.

The original Kalman-Bucy filter was derived for time-varying, linear processes and observations. It may be generalized to encompass nonlinear observations as in the present case where positions and their derivatives are estimated from centrally projected images. In the *extended Kalman filter*, the observation equations are linearized about predicted or *a priori* state-estimates from previous observations, after which the current observations are processed [11, 16].

In a model study, this extended Kalman filter has been investigated for

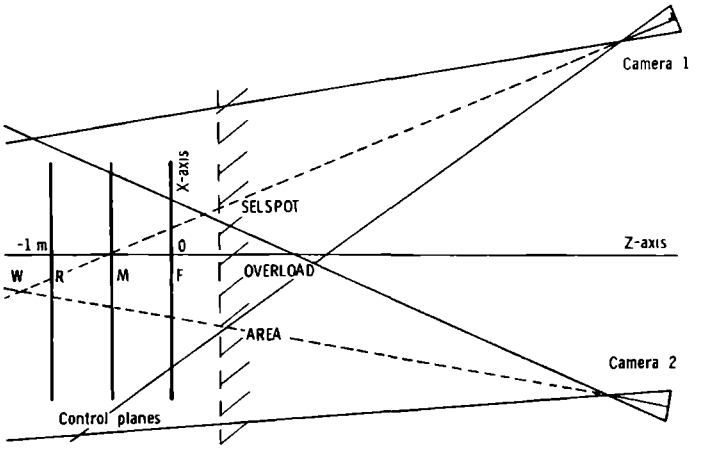


Fig. 3. Control and camera layout in dynamic measurement.

two distortion-free cameras and for a second order model (i.e., two derivatives) with simultaneous estimation of position, velocity and acceleration in 3-D. The results were numerically promising, but the filter was far too slow for real-time data-processing or acceptable analysis delays in biomechanical applications, since the processing time for a single observation from the two cameras amounted to approximately 2 sec (Fortran programming, RT11-F/B operating system V01-11, PDP-11/45 computer, no floating-point hardware). Unlike the Kalman filter parameters in the linear filtering case (Appendix A15–A20), these parameters contain observational information and cannot be stored as tabular data in order to increase calculation speed. For the same reason, the filter will never come in the steady state except for the trivial case of movement termination.

For this reason, some experiments have been run on linear, time-invariant and unidimensional Kalman filtering applied to reconstructed object space coordinates as discussed in the previous section. This is a suboptimal procedure since the state-covariance matrices in the second order model are assumed valid for all coordinates, whereas depth accuracy is actually much smaller than accuracy in the X-Y plane. Observation noise standard deviation SDZ and process position noise standard deviation SDX were considered constant, unlike the position-dependent errors in the raw observations as apparent from table III. For the state-position terms, only their ratio is relevant; however, decision criteria on observation rejection and on filter

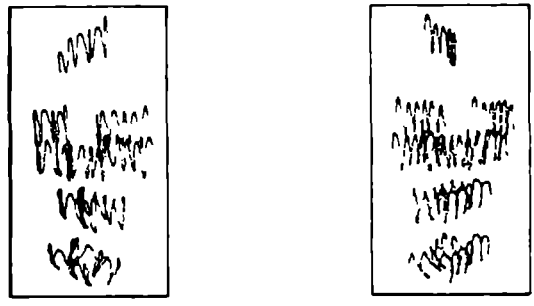


Fig. 4. Raw camera observations on human gait.

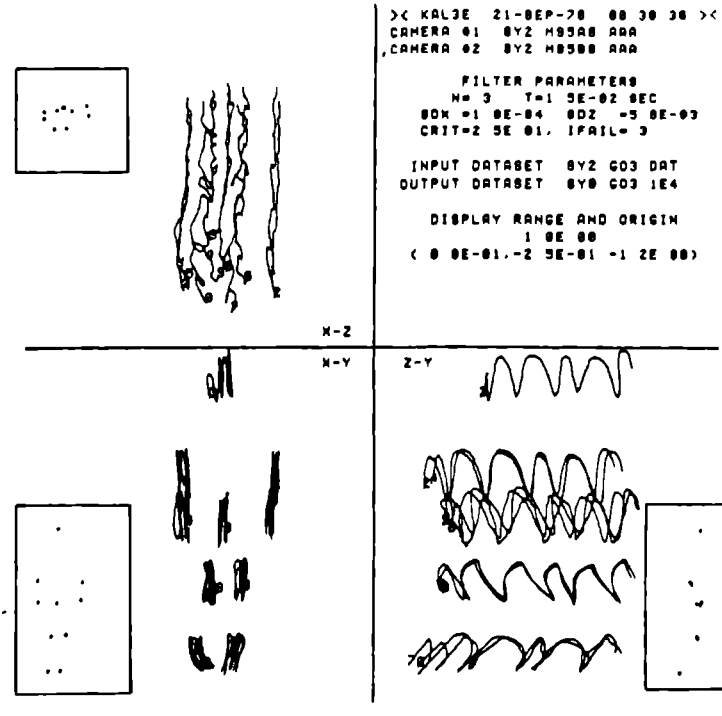


Fig. 5. Reconstructed and filtered trajectories of human gait, projected into the three coordinate planes.

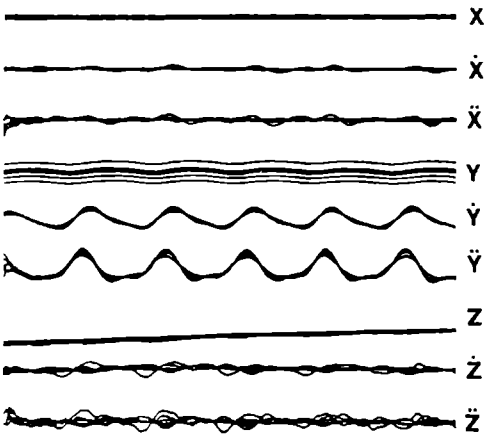


Fig. 6. Estimated positions, velocities and accelerations of some LED markers as a function of time.

reinitialization in the case of complete trajectory loss require absolute values, as discussed below.

As an example, the figures 4–6 show raw camera observations, estimated positions for a single time instant and for a complete measurement interval, and the 9-elements state-vector as a function of time. In this experiment, the subject was slowly hopping toward the cameras, with the LED markers mounted in such a way as to minimize the risk of data loss due to shadowing or optical misalignment. The trajectories and dot patterns for the filtered data are displayed as parallel projections into the object-space coordinate planes, as indicated in figure 5; the photogrammetric reconstruction procedure for the raw position data was based on the explicit, one shot procedure and image distortion model discussed before.

Since these data were filtered rather than smoothed, the state-vector elements suffered from phase-shifts once stronger filtering procedures attained the steady state, beginning with the highest derivatives. As an example, the trajectories in figure 7 show frequent filter *divergence* for the same raw observations as above. Additional research is required to investigate this effect, and to assess the utility of slight fixed-lag smoothing if minor delays in real-time filtering are acceptable. For nonreal-time applications, fixed-interval smoothing is the appropriate procedure [18, 28].

Once the filter attains the steady state, it is an ordinary recursive and constant filter which might have been based on frequency-domain considera-

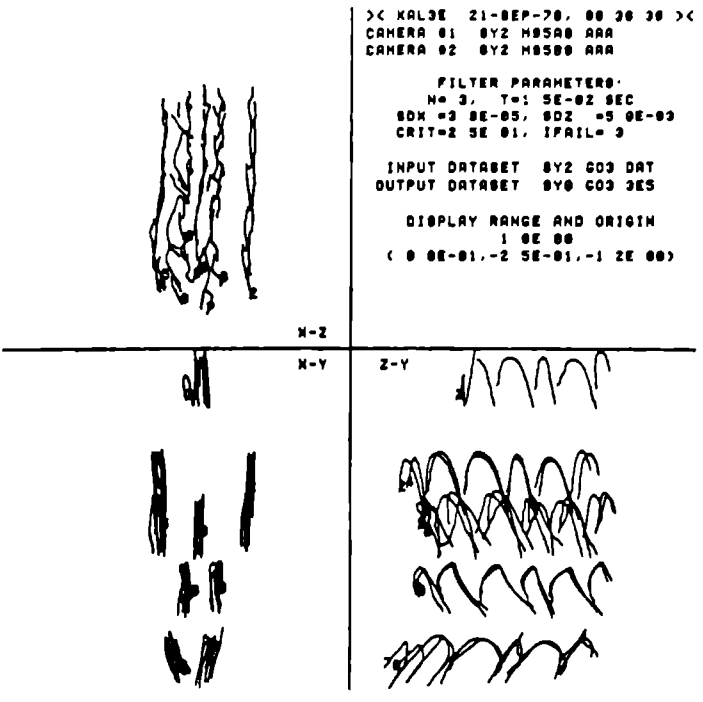


Fig. 7. Filter divergence due to the assumption of little process noise.

tions (A22). This correspondence may guide the choice of the sampling interval T, by estimating the bandwidth over a properly windowed observation interval [33].

The occurence of observation interrupts was handled through bypassing the measurement-update equations (A17–A20); only the time-update equations (A15–A16) were precessed until new observations recurred. However, due to the limited validity of the 2nd order process model, this prediction could not proceed indefinitely, and a failure count was used to suspend filtering until reliable observations recurred. When filtering was suspended such as shown in figure 7, filtering was resumed once two adjacent observations in the raw position data were closer to each other than a criterion value CRIT times the observation-noise standard deviation SDZ. More research is necessary to investigate the use of higher order models for improved process extrapolation during observation interrupts.

The decision to accept or to reject observations is rather difficult. Although the SELSPOT hardware generates overload and out-of-range signals, the

latter are rather unreliable: actual camera output signals become very noisy long before the intercepted light intensity drops to the level corresponding to this binary flag, due to the relative increase of the lateral detector's noise currents [48]. In the present case, observation reliability for allegedly valid SELSPOT data was asserted by comparing the predicted state-position standard deviation (A16) with the innovations position term (A18), i.e. the distance between the 1-step ahead predicted and observed positions. If their ratio was larger than the criterion value CRIT, the observation was rejected, and the failure counter increased. Filtering was suspended as discussed above once a failure count IFAIL was reached. This count was decreased for each accepted observation, with a bottom value zero. If the process position noise standard deviation SDX is chosen too small, this strategy will reinitialize the filter once it diverges too far, as in figure 7.

In the present software, the state-covariance update-equations (A16) and (A19) are bypassed once the filter attains the steady state, with a considerable decrease in calculation time. Single observation cycli have been found to require 100–200 msec for the steady state, 2nd order filter using Fortran programming as indicated above. Filtering speed in the transient case may be increased by storing the Kalman filter parameters as tabular data for the whole transition interval between filter initialization and steady state; in the case of observation interrupts prior to trajectory loss, the covariance time-update equations (A16) may be processed instead of the measurement state-vector update equations (A17) until observations recur. Subsequently, one may choose those parameters in the Kalman filter table which most closely correspond with the currently predicted state-covariance matrix.

In adaptive filtering, the innovations vector (A18) may be used to adjust the process noise parameters [2, 16] or even the filter order. The filtering procedure thus becomes more flexible, at the cost of increased complexity and calculation time requirements; for instance, the filter must now decide whether to reject an observation or to adjust the process model.

Such software decision procedures would be facilitated if the SELSPOT out-of-range flag would exhibit some hysterises for each LED. The position-integrating properties of the lateral photoeffect [47–49] make it impossible to detect false position information as caused by internal lens reflections or by external reflections via neighbouring surfaces in the experimental environment, when an LED moves or is out of view; however, such effects tend to entail large variations in the intercepted light intensity, and the out-of-range signal might shift the criterion of data acceptance to a more conservative level once data rejection has been decided.

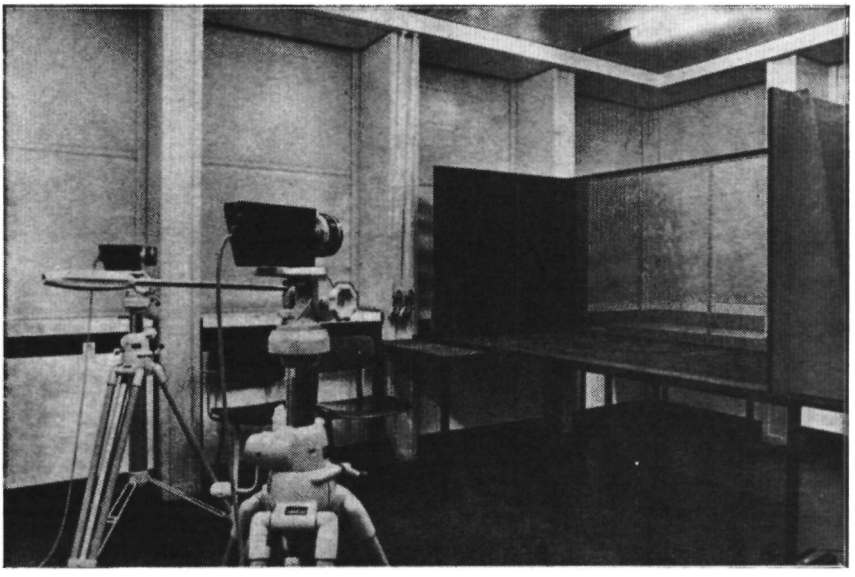


Fig. 8. Calibration object used for camera parameter estimation and absolute accuracy evaluation in static measurement.

As long as the LEDs are within range and oriented toward the cameras, errors due to reflections tend to be negligible unless an LED is close to a polished surface. Such effects were hardly observable in the calibration object shown in figure 8 which was painted with ordinary blackboard paint, although blackness in the visual spectrum does not, of course, imply the absense of reflections in the near infrared as emitted by the LED markers. In general, the use of experimental environments containing unpolished surfaces must be recommended: incident light will then be diffused to all directions rather than reflected into some particular direction, which may coincide with a camera.

Distortion Calibration by Bilinear Polynomial Approximation

The image distortion model in the previous sections was based on physical considerations, containing polynomial elements in the image radius for radial lens distortion, and a term for differential amplification in x and y

amplifiers. An alternative approach is based on general, abstract models not rooted in physical considerations. Since the sole purpose of these models is to compensate for all errors, and not only for all *a priori* envisaged errors, such general models may be preferable: any compound model for a class of envisaged errors may contain highly or even perfectly correlated parameters rendering the corresponding normal or triangularized equation system numerically ill-conditioned or singular; such adverse effects are easier to trace in simple, general models of an abstract nature.

Polynomial approximations are rather common in image calibration. A general model for separate x and y correction may be expressed as

$$f_x(x, y) = a_x^T f(x, y),\ f_y(x, y) = a_y^T f(x, y)\ ^2 \tag{1}$$

where

$$f(x, y) = (1, x, y, x^2, xy, y^2, x^3, \ldots)^T$$

$f(x, y)$ is the general set of polynomial approximation functions, and a_x and a_y are the parameters to be estimated in a least-squares sence.

However, for high polynomial orders, the corresponding condition matrix elements vary over a large range, and this renders the normal or triangularized equation system very ill-conditioned, as manifested by the high values of the so-called *condition numbers* of these matrices, i.e., the ratios of their absolute largest and smallest characteristic values [50]. For a singular matrix, the latter characteristic value is equal to zero.

Based on the work of RAUHALA [37, 38], it is shown in WOLTRING [50] that for approximation functions which are *decomposable* into functions of only one independent variable,

$$f(x, y) = [f_x^T(x) \cdot f_{y1}(y), \ldots, f_x^T(x) \cdot f_{yn_y}(y)]^T \tag{2}$$

with

$$f_x(x) = [f_{x1}(x), \ldots, f_{xn_x}(x)]^T$$
$$f_y(y) = [f_{y1}(y), \ldots, f_{yn_y}(y)]^T$$

and for rectangular control point distributions

$$(x_i, y_j),\ i = 1, \ldots, m_x;\ j = 1, \ldots, m_y \tag{3}$$

the usually overdetermined equation system

$$F\,a = b,\ F^{(m_x m_y) x (n_x n_y)} \tag{4}$$

[2] The following notational conventions are adopted: small letters denote scalars; *italic letters* denote vectors; capitals denote matrices (except for the sampling interval T).

with least-squares solution

$$\hat{a} = F_L^{-1}\, b \tag{5}$$

where

$F_L^{-1} = (F^T F)^{-1} F^T$ the *left inverse* of F

may be written in *matrix* form

$$X\, A\, Y^T = B \tag{6}$$

with

$X = [f_x(x_1), \ldots, f_x(x_{m_x})]^T$
$Y = [f_y(y_1), \ldots, f_y(y_{m_y})]^T$

and where a and b are recovered by adjoining the columns of the parameter matrix A and observation matrix B to single vectors, for increasing column index. The least-squares estimate $\hat{A}$ then follows as

$$\hat{A} = X_L^{-1}\, B\, Y_L^{-T}. \tag{7}$$

This *structured* least-squares procedure involves the inversion of two matrices with dimensions $n_x \times n_x$ and $n_y \times n_y$, respectively, instead of the inversion of an $n_x n_y \times n_x n_y$ matrix as would be required when solving the conventional *monolinear* problem (4) through (5). In addition, memory requirements and calculation time are reduced considerably in this *bilinear least-squares* method. The product of the condition numbers of $X^T X$ and $Y^T Y$ is equal to the condition number of the corresponding normal data-matrix $F^T F$ in (5); consequently, the bilinear least-squares method may be expected to be numerically superior to the conventional monolinear approach.

Additional improvements in numerical stability have been achieved by combining bilinear approximation with orthogonalization procedures as discussed in the Appendix of Part I: some numerical comparisons on SELSPOT distortion observations are presented in [50].

Using bilinear polynomial models, SELSPOT image distortion has been calibrated for rather complete image coverage: 85% in the y-axis, and 90% in the x-axis. The cameras were placed at approximately 2 m from the control plane, with image and calibration axes oriented approximately in parallel. Lens distance setting was 3 m (IR mark), and aperture was taken as 0.95. The data in table IV show image residual standard deviation per axis as a function of the polynomial order, and a 7×7 order polynomial model

Table IV. Residual standard deviations per image axis in bilinear polynomial distortion calibration

17 × 19 control points		Standard deviations per axis (LSB)			
polynomial order	number of parameters	camera No. 1		camera No. 2	
		$\hat{\sigma}_x$	$\hat{\sigma}_y$	$\hat{\sigma}_x$	$\hat{\sigma}_y$
1	4	4.50	4.54	6.03	6.94
2	9	3.67	4.07	4.51	2.66
3	16	1.24	0.97	0.82	0.88
4	25	0.85	0.88	0.82	0.80
5	36	0.72	0.79	0.69	0.72
6	49	0.61	0.70	0.67	0.39
7	64	0.39	0.37	0.35	0.30
8	81	0.35	0.35	0.35	0.25
9	100	0.35	0.35	0.29	0.23

(highest power x^7y^7, 64 parameters) allowed image residual error reduction to almost the quantization level. Calculation times for this distortion calibration procedure are very short, in the order of a few seconds as compared to 5–15 min in the case of the iterated calibration program for joint estimation of internal and external camera parameters discussed above.

Acquisition of the dense and regular observation distribution required for highly accurate distortion calibration may be a rather lengthy process, especially if a single LED is manually moved between control locations as in the present case; this would be hardly worthwhile if system instability necessitates frequent recalibration.

A test of short-term and long-term system stability was carried out as follows: the cameras were placed as above, and not moved with respect to the calibration plane for 3 days; three series of observations were taken three times each, with 24-hour intervals and 2 h equipment warming-up. Parameters were subsequently estimated for one observation set, with varying polynomial orders, and residual standard deviations were calculated for all observation sets. In all possible combinations, the residual standard deviations for the other observation sets remained within 130% of the residual standard deviation of the reference set. As before, temporal noise was rendered negligible by averaging some 330 observations per control point.

These results suggest that improved spatial accuracy in 3-D reconstruction with strong nonphotogrammetric lenses is feasible. A problem, though, is the need of a regular distribution in the independent variables: in the case of extended Kalman filtering, a regular calibration grid in object space is appropriate since undistorted image points corresponding to predicted object-space positions serve as independent variables; they are transformed to distorted predictions through the bilinear model in order to assess the innovations vector in the measurement update-equations (A18). However, for the simplified, linear and time-invariant Kalman filter of the previous section, unfiltered object space coordinates would require distortion calibration in the reverse form, with the *distorted* image patterned in rectangular form. Some methods are suggested [50] for the rectification or *translocation* to such a regular form of distorted images obtained from a regular grid in object space.

Discussion

Calibration by recourse to an absolute control distribution is rather elaborate or expensive during acquisition, especially for image distortion calibration, which requires a large number of observations for acceptable modelling. In the present case, depth errors due to pressure deformation in the perforated, partially reinforced control plane could be limited to ± 0.5 mm, which is substantially below the depth errors in static and dynamic reconstruction as presented in tables II and III. Level errors in the experimental room and positioning errors in the three control plane positions are expected to be of the same order; positioning errors within the control plane have tolerances below 0.1 mm. However, it is unknown which apparent position shifts are caused by refractions in the calibration LED-to-air boundary; since this boundary has an approximately spherical shape (Texas Instruments TIXL 27), and since the actual emitting area has a radius of 0.1 mm, the effect is expected to be negligible.

The high costs of IR power-LEDs for use at larger distances are currently prohibitive for a 3-D, dense control object not requiring manual LED positioning. However, the bilinear stability tests suggest that image deformation may be calibrated at intervals of at least a few days without significant accuracy losses, and the 9 parameters of the distortion-compensated camera – at observation distances which more or less correspond with those during distortion calibration – may be estimated by recourse to a sparse

control object. Such an object is currently under construction in our laboratory, containing 13 LEDs at the apexes and in the rear plane of a 1 m^3 cube to be used for calibration in arm motion studies.

A similar approach in which control point locations are assumed error-free has been adopted by TORLEGÅRD [45]. He employed photogrammetric cameras and showed that precision and independence of parameters improve with increasing relative spatial extension of a 3-D control distribution. Focal length and control-to-camera distance become increasingly separated for increasing control depth with respect to the camera; similarly, the interdependence of camera attitude φ_C and principal point location x_0 (cf. table I in Part I) diminishes with increasing viewing angle [20, 45]. Low correlations between parameters allow accurate reconstruction outside the area delimited by the original control distribution *(extrapolation)*, rather than only inside this distribution (strict *interpolation*).

The cumbersomeness of absolute, 3-D control may be avoided by carrying the positions and attitudes of the physical control planes as adjustable parameters into the condition equations *(partial control)*. In this method of *simultaneous multiframe analytical calibration (SMAC)* [8], a number of oblique observations on a 2-D absolute control distribution (at least 3 frames) for different *swing* angles about the optical camera axis warrant sufficient determinacy for a complete recovery of internal and external camera parameters. In essence, the obliqueness of the observation disrupts the perfect correlation between focal length and observation distance, while the use of different swing angles will adjoin correlated and anticorrelated observations for principal point location and camera attitude.

More recently, recovery of internal parameters has been found possible without recourse to absolute control in object space [8, 20]. By viewing a 3-D distribution of control points with unknown locations *(relative control)* in a similar way as above, and by assuming an arbitrary camera station for one observed frame, and an arbitrary scale factor in object space, a stable solution of the internal camera parameters, including focal length, may be obtained. External parameters become recoverable by defining a coordinate system on, for instance, three noncollinear targets. Thus, one might distribute the SELSPOT LEDs over the future observation area, measure the distance between two LEDs and indicate a third LED for the object-space frame of reference, and observe the LEDs from various viewpoints. The cameras are left in the last positions and attitudes, and the calibration program may be run while the subject is being equipped with the LED body markers. Sparseness of the condition matrix F_x (cf. Part I) in iterative parameter

adjustment may render these calibration procedures relatively efficient, although the increased number of unknowns as compared to absolute control methods has a considerable influence on computation time. BROWN [8] has presented an extensive review of the various methods.

In iterative adjustment, an initial estimate must be given. Some parameters such as for image distortion may be initialized to zero, others must be guessed by visual inspection or other nonphotogrammetric methods. In the case of relative control, the unknown control point locations may be initially estimated by carrying out explicit space intersection on the basis of initially estimated camera parameters. In the case of *absolute* control, the *direct linear transformation* [1] allows another initialization approach, under the assumption of distortion-free cameras. Here, an explicit least-squares solution is given by neglecting the proportionality factor k in the projective equations (3) of Part I and the orthogonality of the camera image and optical axes.

Reconstruction following the above calibration procedures is basically a multidimensional approximation method in which the approximating functions are (partially) based on a physical model of the camera observation process. In a similar way as for the bilinear approximation method for image distortion compensation, one may wonder whether general approximation functions of a more abstract nature might be used instead. The work of RAUHALA [37, 38], which motivated the use of bilinear approximation is very promising in this respect: bilinear approximation may be easily extended to the n-dimensional case, and by using regular n-dimensional (absolute) control distributions one may achieve savings in computation time and memory requirements that allow parameter estimation problems to be solved which are unsurmountable when using conventional monolinear least-squares. For such approximation models, there is no need to assume distortion invariance with observation distance changes, and other variables such as light intensity in the case of improper SELSPOT signal normalization might be included. However, such models may require more parameters than would be necessary for physically oriented models, and the computational load on currently available, serially operating computers may become prohibitive during measurement. The increased interest in parallel processors (Array computers, Fast Fourier Transform hardware) is quite promising in this respect, especially because of the regular structure of multidimensional matrix or *array* algebra.

Similar remarks pertain to time-domain filtering. The estimation of multidimensional parameters along RAUHALA's lines becomes efficient by

factorizing approximation functions and control distributions in terms of the individual independent variables, and these may be processed in parallel. Similarly, LAINIOTIS [23] partitions a measuring interval into a set of subintervals which are completely decoupled for adaptive estimation of state variables; these subintervals thus become amenable to parallel processing with again considerable savings in computation time and memory requirements.

Alternatively, hybrid computation techniques are very interesting; for instance, nonlinear observation equations might be solved in digital form, and the calculated Kalman gain factors and innovations (A18, A20) may be fed into an analog computer – which is a parallel implementation for continuous signals.

Such facilities are quite costly, and some funds might be used to improve the movement monitoring hardware itself rather than to correct its limitations in subsequent data processing. For instance, the use of photogrammetric lenses and improvement of SELSPOT out-of-range signals have already been discussed. Also, very strong AC background light of an incandescent nature (e.g., floodlighting in the case of simultaneous filming) may interfere with the present signal-background light compensation process, and an improved SELSPOT system will counter this effect, by averaging the lateral detector's output currents before and after an LED light pulse, rather than relying only on the detector currents prior to a light pulse as in the present situation [Selcom AB, pers. commun.].

The possibility of optical SELSPOT synchronization allowing the subject to move freely from the cameras is a very interesting facility; unfortunately, it relies on a single LED which emits a longer light pulse than the other LED targets do, and prolonged observation loss of this LED during, say, a few seconds, may cause complete system failure. A better approach might be to use a phase-lock loop correlating adjacent observation frames of 1–30 LEDs, which adjusts the receiver clock if correlation changes are detected in the pattern as formed by all operating LEDs – assuming rather small dynamic changes in the motion pattern between adjacent frames.

Despite these minor limitations, use of the SELSPOT system appears to be an interesting approach for trajectory reconstruction as compared to traditional film analysis. At the same time, however, the limited temporal resolution as apparent from the results in table III has considerable influence on wide-band differentiation for velocity and acceleration estimation which are necessery in the investigation of the microstructure of human motion. Some improvements may be expected from the use of currently available

lateral detectors with a lower noise level or from still stronger lenses and LEDs, for a certain observation distance; the latter two, however, would entail increased image distortion and heating of the LED light sources on the subject's body, respectively.

It is of some interest to note in this context, that MITCHELSON [30, 31] has described a 12–14 bits optoelectronic system based in part on the binary Gray-coded detector as previously discussed [47]. This *CODA system* (for Cartesian optoelectronic dynamic anthropometer) is now operational with 12 bits resolution per axis, at 1 % full-scale lens errors which at the present time are not yet compensated. Sampling frequency is 1 kHz, and up to eight laser diodes may be accomodated as body markers [MITCHELSON, pers. commun.].

In summary, one may expect that real-time monitoring of human motion at the level of simultaneous estimation of 3-D positions and time derivatives for multiple targets will be feasible within a few years. This will allow the introduction of efficient training procedures in sports, by giving subjects immediate feedback about their movement trajectories as compared to reference trajectories, assuming that such reference trajectories can be constructed, e. g. from well-trained performers of a similar corporal structure. Alternatively, present delays in gait-data processing for clinical purposes may be rendered negligible through such developments.

Even in those applications where ultimate interest is focused on 2-D projections of motion, as for instance in certain pattern recognition studies in psychology [17, 26], the use of 3-D data acquisition and processing schemes allows flexibility as regards the projection direction, compensation of parallax, and reduction of the risk of data loss due to shadowing effects. Thus, the side view of the gait pattern in figure 5 would have been difficult to obtain by recording the subject's movements laterally as is often done in gait studies.

The requirement of wired body markers as in the SELSPOT system and in MITCHELSON's CODA system may be quite acceptable for certain constrained movement types such as in arm movement studies or standardized gait measurement. However, in those cases where hindrance of the subject must be minimized, such as in golfing studies and in gait analysis with children, they present a serious drawback. It is to be expected that current developments in optoelectronics and in parallel signal processing will allow real-time implementation of systems requiring only passive targets in conjunction with 3-D identification at the receiving end [47] within the next few years.

Appendix: a summary of Kalman Filtering

State of a Process

Contemporary time-domain description of a process (or system) evolves around the notion of the *state* of the process under study, i.e. the behavior of a combination of those current aspects of the process which, together with future process inputs, completely specify future behavior of the process, or are assumed to do so. An example of a *state vector* is the position and velocity of a vehicle at any time instant, with accelerations or acting forces as input variables.

The state is a useful concept for *causal* processes [21], that is, where future inputs have no influence on present and past process behavior. One should note that this is different from *estimating* current state elements from future behavior as in smoothing, which is based on the assumption of correlations between process aspects for neighboring time instants.

For continuous processes, the state may be denoted as a vector-valued function of time $x(t)$; for discrete processes or discrete observations on continuous processes, the state vector becomes a function x_k of a discrete time-index k. The intervals may be equidistant with sampling interval T as assumed below, but do not have to be so.

Continuous Process Description

Time-domain description of linear, constant and continuous processes centers around the matrix notation of a first order, linear equation system with *process* or *system matrix* F,

$$\dot{x}(t) = F\,x(t) + B\,u(t) \tag{A1}$$

with $u(t)$ the input to the process at time t, distributed over the differentiated state vector $\dot{x}(t)$ through the *input* or *associated matrix* B. The observable process aspects such as position in the present case are considered to be linear functions of the state vector, i.e.,

$$z(t) = H\,x(t). \tag{A2}$$

Generalization to time-varying, nonlinear processes and observations is possible, and is denoted as

$$\begin{aligned} \dot{x}(t) &= f[x(t),u(t),t] \\ z(t) &= h[x(t),t] \end{aligned} \tag{A3}$$

Assuming zero initial conditions, formation of the Laplace transform of relations A1 and A2, and elimination of the state vector result in the joint process-measurement transfer function

$$G(s) = H\ (sI - F)^{-1}\,B \tag{A4}$$

in which each element in the matrix G(s) specifies the transfer function of the corresponding input-output combination. In this manner, a link is made with conventional frequency-domain analysis.

In the present case, where interest is directed to simultaneous estimation of positions and their derivatives, the state vector for the unidimensional case is defined as

$$x(t) = [x(t), \dot{x}(t), \ddot{x}(t), \ldots]^T \tag{A5}$$

with corresponding system matrix

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ . & & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix} \tag{A6}$$

For an n-th order process, F has dimensions $(n+1)\times(n+1)$.

In the Kalman filtering context, A1 is a local approximation of the observed process, with optimally estimated model parameter vector $x(t)$, and with the transformed input vector $u(t)$ a white noise term with known covariance matrix. For the present case, B $u(t)$ may be assumed to affected the highest element in the differentiated state vector only, since the system matrix part of the model assumes that the highest derivative in the state vector is constant, which may or may not be true in reality. Thus, taking $B=I$,

$$u(t)=[0, 0, 0, ..., u(t)]^T \tag{A7}$$

with:

$$E[u(t)]=0,\ E[u(t)\cdot u(s)]=\sigma_n^2\delta(t-s)$$

where $\delta(t-s)$ is the *delta function.*

If the modelling error appears correlated over a prolonged time interval, divergence may occur, and one may extend the state vector with *bias* parameters to remove the colored noise components.

Discrete Time Description

If the process description A1-5-6-7 is required only at discrete time-instants, e.g. as a result of intermittent observations with a sampling interval T,

$$z_k = Hx_k \tag{A8}$$

A1 may be integrated between observations [14] ro render

$$x_k = \Phi x_{k-1} + v_k \tag{A9}$$

with:

$\Phi = \Phi(T) = e^{FT}$, the *transition matrix* of the equivalent discrete process

$$v_k = \int_{t_{k-1}}^{t_k} e^{F(t_k-\tau)} B\, u(\tau)\, d\tau,$$ discrete input signal version.

For the *exponential matrix* $e^{F\tau}$, the reader is referred to the literature, e.g. [14]. In essence, $e^{F\tau}$ may be expressed as a Taylor series of matrix products

$$e^{F\tau} = \sum_{i=0}^{\infty} \frac{1}{i!} F^i \tau^i,$$

where powers of F for $i>n$ may be calculated through characteristic value analysis from the powers of F for $i\leqslant n$.

If the continuous input $u(t)$ may be considered constant between observations, the discrete process input vector u_k becomes

$$u_k = G\, u(t_k) \tag{A10}$$

with:

$$G = \int_0^T e^{F\tau}\, B\, d\tau.$$

For general, time-varying $u(t)$ between observations, no such simple relation exists.

The transition matrix corresponding to the process matrix (A6) is known to be the generalized *Newtonian* matrix:

$$\Phi(T) = \begin{bmatrix} 1 & T & \tfrac{1}{2}T^2 & \dots \\ 0 & 1 & T & \dots \\ 0 & 0 & 1 & \dots \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{bmatrix} = [\varphi_{i,\,i+j}],\ \varphi_{i,\,i+j} = \begin{cases} 0 & , j<0 \\ \dfrac{1}{j!}T^j, & j\geqslant 0 \end{cases} \tag{A11}$$

and the discrete, white noise vector v_k corresponding to (A7) has zero mean and covariance matrix

$$\operatorname{cov}(v_k) = E\{v_k v_l^T\} = Q\,\delta_{kl}$$

with:

$$Q = [q_{i,j}]$$

$$q_{n+1-i,\,n+1-j} = \sigma_n^2 \int_0^T \frac{\tau^i}{i!}\frac{\tau^j}{j!}\, d\tau = \frac{\sigma_n^2\ T^{i+j+1}}{i!\,j!\,(i+j+1)},\ 0\leqslant i,j\leqslant n \tag{A12}$$

as may be shown by generalizing SINGER's [39] results for white noise in a 2nd order model.

If in addition the observations are disturbed with zero mean, white noise, with covariance matrix $R\,\delta_{ij}$ (δ_{ij} is the Kronecker delta), the complete discrete-time process and measurement description becomes:

$$x_k = \Phi\, x_{k-1} + v_k,\ v_k \text{ white, zero-mean noise with cov. matrix } Q\delta_{kj} \tag{A13}$$

$$z_k = H\, x_k + w_k,\ w_k \text{ white, zero-mean noise with cov. matrix } R\delta_{kj}, \tag{A14}$$

v_k and w_k are usually considered uncorrelated, as in the present case.

The equations A2, A8 and A14 are usually called *observation* equations, and correspond to the condition equations in the photogrammetric terminology of Part I.

Discrete Kalman Filtering

The Kalman filter starts from the process and measurement description A13 and A14, and assumes in addition that x_k is a noisy variable with initial estimate x_o and initial covariance matrix estimate P_o; the state estimate is supposed to be uncorrelated with v_k and w_k. The recursive and linear, optimal Kalman filter combining the state value x_k as

predicted from past observations with current observations in a minimum variance sense is known to be [6, 11, 16]:

1. Time Update (1-Step Prediction)

$$\bar{x}_k = \Phi \hat{x}_{k-1} \quad \text{predicted state} \tag{A15}$$

$$\bar{P}_k = \Phi \hat{P}_{k-1} \Phi^T + Q \quad \text{predicted state covariance matrix} \tag{A16}$$

2. Measurement Update (Filtering)

$$\hat{x}_k = \bar{x}_k + K_k y_k \quad \text{filtered state} \tag{A17}$$

$$y_k = z_k - H \bar{x}_k \quad \text{innovations vector} \tag{A18}$$

$$\hat{P}_k = \bar{P}_k - K_k H \bar{P}_k \quad \text{filtered state covariance matrix} \tag{A19}$$

$$K_k = \bar{P}_k H^T (H \bar{P}_k H^T + R)^{-1} \quad \text{Kalman gain} \tag{A20}$$

with: $P_k = \text{cov}(x_k)$; $^-$ denotes predicted values from past observations (*a priori* information); $\hat{}$ denotes filtered values including current observations (*a posteriori* information).

Steady State Kalman Filter

By eliminating the 1-step predicted state vector $\bar{x}_k$ from A15 and A17, one finds

$$\hat{x}_k = (I - K_k H) \Phi \hat{x}_{k-1} + K_k z_k. \tag{A21}$$

Thus, the Kalman filter is expressed as a recursive filter of a time-varying nature because of the evolution of the Kalman gain through the covariance update equations A16 and A19. At each time instant, predicted state and incoming information are weighted in accordance with their respective covariance matrices $\bar{P}_k$ and R. Since the covariance update equations for the linear filter do not contain the observations, the P_k may be calculated beforehand and stored as tabular data. Unfortunately, no explicit solution as a function of the time index k can be given for the gain in general; however, K_k is known to converge to a steady value for the types of transition matrices and noise covariance matrices discussed in this Appendix; the filter becomes an ordinary constant filter which may be described in frequency terms via the z-transform and Laplace transform [11, 14]:

$$X(s) = G(s) \cdot Z(s) \tag{A22}$$

with:

$$G(s) = [I - e^{-sT}(I - K_\infty H) \Phi]^{-1} K_\infty.$$

Factorization Methods

The covariance-matrix measurement-update formula A19 is known to be numerically unreliable since the difference between two semidefinite positive matrices is calculated, which may result in filtered state-covariance matrices with negative variance elements on their main diagonals or in off-diagonal correlation terms greater than unity. A better approach is to employ certain factorizations of P_k in a similar vein as discussed in the Appendix of Part I where normal datamatrices were avoided by triangularization procedures. In the present case, BIERMAN's U-D factorization [5, 44] has been used, which has appeared to be virtually equal to GENTLEMAN's [13] version of Givens rotations discussed in Part I. Whereas GENTLEMAN's procedure corresponds to factoring a symmetrical matrix P into a diagonal matrix D and upper unit triangular matrix R, with $P = R^T D R$,

BIERMAN's U-D factorization has upper unit triangular U and diagonal D in $P = U\,D\,U^T$. By employing such factorization methods, the covariance-update equations A16 and A19 are replaced by update equations on the U-D factors, with some savings in calculation time. In addition, the sparse structure of various matrices may be easily exploited to speed up the calculation process even more. A review of various other methods is available in BIERMAN [6].

Parameter Choice

For the linear, 2nd order filter used in the present experiments, the transition matrix followed from A11, with observation matrix $H = (1, 0, 0)$, for noisy position observations assumed independent in the three object-space coordinates.

The process-noise covariance matrix A12 was normalized with respect to the position standard deviation, whose value SDX served as filter input parameter, together with the observation noise standard deviation SDZ ($SDZ^2 = R$).

The filter was initialized by using two adjacent observations to initialize the state position and velocity; acceleration was simply set to zero. The n-th order state-covariance matrix was initialized by assuming that all derivatives were obtained by finite differences of uncorrelated observations, each with standard deviation SDZ. One may prove that P_o then adopts the form:

$$P_o = L\,D\,L^T \tag{A23}$$

with:

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & -1 & 0 & 0 & \cdots \\ 1 & -2 & 1 & 0 & \cdots \\ 1 & -3 & 3 & -1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot \end{bmatrix}$$ Pascal's inverse triangular matrix,

$$D = \mathrm{diag}(T^{-i}) \cdot SDZ^2, \quad i = 0, 1, \ldots, n.$$

Acknowledgements

The author would like to thank Mr. H. WILLEMSE for the construction of the 3-D control object shown in figure 8, and Mr. J. WITTEBROOD for the construction of an interface for remote SELSPOT data transmission in direct memory access to a PDP-11/45 computer. In addition, he is grateful to Prof. Dr. Ir. E. G. J. EIJKMAN for valuable comments on Part I of this paper, and to Dr. G. J. BIERMAN of Jet Propulsion Laboratories, Pasadena, Calif., for stimulating discussions on current developments in time-domain filtering, during a recent study-voyage in the United States of America and Canada which was made possible in part through a travel grant from the Netherlands Organization for the Advancement of Pure Research (ZWO). Finally, he is indebted to the Department of

Photography of the University of Nijmegen for the processing of the illustrations, and to Mrs. J. A. THOMASSEN-BAKER for linguistical suggestions on both manuscripts.

References

1 ABDEL-AZIZ, Y. I. and KARARA, H. M.: Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry; in Proc. of the Symp. on Close-Range Photogrammetry, Urbana, Ill. 1971, pp. 1–18 (American Society of Photogrammetry, 1971).
2 ALDRICH, G. T. and KRABILL, W. B.: An application of Kalman techniques to aircraft and missile radar tracking. AIAA J. *11:* 932–938 (1973).
3 ALSPACH, D. L.: A gaussian sum approach to the multi-target identification-tracking problem. Automatica *11:* 285–296 (1975).
4 BECKER, E. B.: Measurement of mass distribution parameters of anatomical segments. Report NAMRL-1193, Naval Aerospace Medical Research Laboratory, Pensacola, Fla. 1973.
5 BIERMAN, G. J.: Measurement updating using the U-D factorization. Proc. of the 1975 IEEE Conf. on Decision and Control, Houston, Tex. 1975, pp. 337–346. Also in Automatica *12:* 375–382 (1976).
6 BIERMAN, G. J.: Factorization methods for discrete sequential estimation (Academic Press, New York 1976).
7 BROOKS, C. B. and JACOBS, A. M.: The gamma mass scanning technique for inertial anthropometric measurement. Med. Sci. Sports *7:* 290–294 (1975).
8 BROWN, D. C.: Evolution, application and potential of the bundle-method of photogrammetric triangulation. Symposium Commission III, Int. Soc. for Photogrammetry, Stuttgart 1974 (DBA Systems Inc., Fla. 1974).
9 CHAO, E. Y. S. and RIM, K.: Dynamic modelling of human lower limbs. ASCE National Structural Engineering Meeting, Cincinatti, Ohio 1974.
10 DRILLIS, R. and CONTINI, R.: Body segment parameters. Tech. Rep. 1166.03, School of Engineering and Science (New York University, New York 1966).
11 EYKHOFF, P.: System identification. Parameter and state estimation (Wiley, New York 1974).
12 GAUSS, K. F.: Theory of the motion of the heavenly bodies moving about the sun in conic sections. (Translation.) (Dover, New York 1963.)
13 GENTLEMAN, W. M.: Least squares computations by Givens transformation without square roots. J. Inst. Maths. Applics. *12:* 329–336 (1973).
14 GUPTA, S. C.: Transform and state variable methods in linear systems (Wiley, New York 1966).
15 HERRON, R. E.; CUZZI, J. R., and HUGG, J.: Mass distribution of the human body using biostereometrics. Report AMRL-TR-75-18, National Technical Information Service, Springfield, Va. 1976.
16 JAZWINSKI, A. H.: Stochastic processes and filtering theory (Academic Press, New York 1970).
17 JOHANSSON, G.: Visual perception of biological motion and its analysis. Perception and Psychophysics *14:* 201–211 (1973).

18 KAILATH, T.: Supplement to Medith's 'A survey of data smoothing'. Automatica *11:* 109–111 (1975).
19 KANYUCK, A. J.: Transient response of tracking filters with randomly interrupted data. IEEE Trans. on Aerospace and Electronic Systems, *6:* 313–323 (1970).
20 KENEFICK, J. F.; GYER, M. S., and HARP, B. F.: Analytical selfcalibration. Photogramm. Engng *38:* 1117–1126 (1972).
21 KWAKERNAAK, H.: Dynamic systems and control; in DALENOORT Process models for psychology (University Press, Rotterdam 1973).
22 KWAKERNAAK, H. and SIVAN, R.: Linear optimal control systems (Wiley, New York 1972).
23 LAINIOTIS, D. G.: Partitioning: a unifying framework for adaptive systems. I. Estimation. (Special issue on adaptive systems.) Proc. IEEE *64:* 1126–1143 (1976).
24 LINDHOLM, L.-E.: An optoelectronic instrument for remote on-line movement monitoring; in NELSON and MOREHOUSE Biomechanics, vol. IV, pp. 510–512 (University Park Press, Baltimore 1974).
25 LEONDES, C. T.: Theory and applications of Kalman filtering. AGARDOgraph 139 (Agard, Paris 1970).
26 LEVELT, W. J. M.; SCHREUDER, R., and HOENKAMP, E. C. M.: Structure and use of verbs of motion; in CAMPBELL and SMITH Proc. of the Conf. on the Psychology of Language, Stirling 1976 (to be published).
27 MCCLELLAN, J. H.; PARKS, T. W., and RABINER, L. R.: A computer program for designing optimal FIR linear phase digital filters. IEEE Trans. Audio Electroac. *21:* 506–526 (1973).
28 MEDITH, J. S.: A survey of data smoothing for linear and nonlinear dynamic systems. Automatica *9:* 151–162 (1973).
29 MILLER, D. J. and NELSON, R. C.: Biomechanics of sport, pp. 245–246 (Lea & Febiger, Philadelphia 1973).
30 MITCHELSON, D. L.: An opto-electronic technique for analysis of angular movements; in CERQUIGLINI *et al.* Medicine and Sport, vol. 8: Biomechanics III (Karger, Basel 1973).
31 MITCHELSON, D. L.: Recording of movement without photography; in GRIEVE *et al.* Techniques for the analysis of human movement (Lepus, London 1975).
32 MORRISON, N.: Introduction to sequential smoothing and prediction (McGraw-Hill, New York 1969).
33 OPPENHEIM, A. V. and SCHAFER, R. W.: Digital signal processing (Prentice-Hall, Englewood Cliffs 1975).
34 PEZZACK, J.; NORMAN, R. N., and WINTER, D. A.: An assessment of derivative determining techniques used for motion analysis. Proc. 11th IFMBE, Ottawa 1976 (to be published).
35 PLAGENHOEF, S.: Patterns of human motion (Prentice-Hall, Englewood Cliffs 1971).
36 RABINER, L. R. *et al.:* Terminology in digital signal processing. IEEE Trans. Audio Electroac. *20:* 322–337 (1972).
37 RAUHALA, U. A.: Array algebra with applications in photogrammetry and geodesy; doctoral thesis, Royal Institute of Technology, Stockholm. Fotogrammetriska Meddelanden 6, No. 6 (1974).
38 RAUHALA, U. A.: A review of array algebra. Fotogrammetriska Meddelanden *2:* 38 (1976).

39 SINGER, R. A.: Estimating optimal tracking filter performance for manned maneuvering targets. IEEE Trans. on Aerospace and Electronic Systems *6:* 473–483 (1970).
40 SINGER, R. A. and BEHNKE, K. W.: Real-time tracking filter evaluation and selection for tactical applications. IEEE Trans. on Aerospace and Electronic Systems *7:* 100–110 (1971).
41 SINGER, R. A. *et al.:* Derivation and evaluation of improved tracking filters for use in dense multitarget environments. IEEE Trans. Info. Th. *20:* 423–432 (1974).
42 SMITH, A. J.: A study of forces on the body in athletic activities with particular reference to jumping; unpubl. PhD. thesis, University of Leeds (1972).
43 STEIN, J. J. and BLACKMAN, S. S.: Generalized correlation of multitarget track data. IEEE Trans. on Aerospace and Electronic Systems *11:* 1207–1217 (1975).
44 THORNTON, C. L. and BIERMAN, G. J.: Gram-Schmidt algorithms for covariance propagation. Proc. of the 1975 IEEE Conf. on Decision and Control, Houston, Tex. 1975, pp. 489–498.
45 TORLEGÅRD, K.: On the determination of interior orientation of close-up cameras under operational conditions using three-dimensional test objects; doctoral thesis, Royal Institute of Technology, Stockholm (1967).
46 WINTER, D. A. *et al.:* Analysis of instantaneous energy of normal gait. J. Biomech. *9:* 253–257 (1976).
47 WOLTRING, H. J.: New possibilities for human motion studies by real-time light spot position measurement. Biotelemetry *1:* 132–146 (1974).
48 WOLTRING, H. J.: Single- and dual-axis lateral photodetectors of rectangular shape. IEEE Trans. Electron. Dev. *22:* 581–590, 1101 (1975).
49 WOLTRING, H. J.: Calibration and measurement in 3-D monitoring of human motion by optoelectronic means. I. Preliminaries and theoretical aspects. Biotelemetry *2:* 169–196 (1975).
50 WOLTRING, H. J.: Bilinear approximation by orthogonal triangularization. Photogrammetria *33:* nr. 3 (1977), (in print).

H. J. WOLTRING, Laboratory of Psychology, University of Nijmegen, Erasmuslaan 16, *NL-6500 HE Nijmegen* (The Netherlands)

Photogrammetria, 33 (1977) 77—93

"Ik ga uit van der existens verscheidener werelden in eenzelfder tijdsverband," sprak hij. "Deze gans verscheidene werelden kunnen zich op een bepaald punt raken, volgt u mij? Kijk, twee maal twee is slechts schijnbaar vier. Men moet rekening houden met der slijtage-fenomeen dat door multipliceren veroorzaakt wordt. En der slijtage-fenomeen kan ook optreden in der door mij geschetste omstanden. Let u thans goed op ...".

2.2

Prof. Prlwytzkofski, (Heer Bommel en) de kwade inblazingen.

BILINEAR APPROXIMATION BY ORTHOGONAL TRIANGULARIZATION

H.J. WOLTRING

Laboratory of Psychology, The University of Nijmegen, Nijmegen (The Netherlands)

(Received October 12, 1976; revised and accepted December 20, 1976)

ABSTRACT

Dimensionwise factorization in global approximation procedures for analytical image distortion calibration has a beneficial effect in terms of numerical stability, calculation speed and memory requirements, at the cost of certain limitations on the class of approximation functions and observations distribution. The formal equivalence of Rauhala's method for dimensionwise factorization to conventional least squares using rectangular calibration grids is proven, and numerical stability is analyzed in terms of the condition numbers of normal datamatrices. Additional improvements appear to be feasible by employing orthogonal matrix triangularization methods rather than normal equations; this is demonstrated with experimental results obtained with an optoelectronic system for infrared light source direction monitoring typically used in biomechanical analysis of human motion.

INTRODUCTION

The precision of analytical reconstruction methods in photogrammetry depends to a great extent on the accuracy of calibrating image distortion. For reasons of economy, image distortion cannot usually be avoided, and this may necessitate the use of efficient analytical procedures in calibration.

Many distortion sources have a global effect on image quality, including: (1) lens errors (radial and tangential distortion); (2) defocusing (shifts of blurred image element *centroids*); (3) film shrinkage; and (4) drift and gain errors in electronic image processors (TV).

A common approach is to define a sufficiently general, parametric model for all errors, and to estimate the parameters of this model in some optimal sense, usually by a least-squares method. In this fashion, one may avoid unintentionally high correlations between parameters in a combined model for all envisaged error sources such as those above.

Yet, numerical problems tend to influence the quality of such general approximation (or interpolation) procedures, especially in the case of polynomial approximation. Therefore, the combination of two methods for improved

numerical accuracy and calculation speed in least-squares parameter estimation is advocated in this paper and demonstrated with some experimental results.

Data were obtained with an optoelectronic system for infrared light source direction monitoring SELSPOT — for SElective Light SPOT recognition (Anonymous, 1975) — which is typically used to measure dynamic processes such as those studied in the field of biomechanics.

First, however, a resumé on linear least squares along traditional lines is presented.

LINEAR LEAST SQUARES

For reasons of mathematical trackability, linear models are usually chosen, i.e., one defines a class of functions in the independent variables x and y (usually the distorted image coordinates):

$$f^{\mathrm{T}}(x,y) = \{f_i(x,y)\}, \qquad i = 1,\ldots,N \tag{1}$$

and tries to find a suitable, linear combination of these functions as defined by a parameter vector a in:

$$f(x,y) = a^{\mathrm{T}} f(x,y), \qquad \text{with: } a = \{a_i\}, \qquad i = 1,\ldots,N \tag{2}$$

where the function $f(x,y)$ should approximate the dependent variable in some optimal sense. Such parameter vectors and approximation functions are sought for both the undistorted x- and y-coordinates: a_x and $f_x(x,y)$, and a_y and $f_y(x,y)$, respectively.

The vectors a_x and a_y may be estimated by recourse to a set of observations:

$$b_x^{\mathrm{T}} = \{b_x(x_i,y_i)\}, \qquad b_y^{\mathrm{T}} = \{b_y(x_i,y_i)\}, \qquad i = 1,\ldots,M \tag{3}$$

which results in two sets of *observation equations*:

$$\left.\begin{aligned} f^{\mathrm{T}}(x_i,y_i)\cdot a_x &= b_x(x_i,y_i) \\ f^{\mathrm{T}}(x_i,y_i)\cdot a_y &= b_y(x_i,y_i) \end{aligned}\right\}, \qquad i = 1,\ldots,M \tag{4}$$

and which may be written in *matrix* form as:

$$\mathbf{F}\cdot(a_x,a_y) = (b_x,b_y), \qquad \text{with: } \mathbf{F} = [f(x_1,y_1),\ldots,f(x_M,y_M)]^{\mathrm{T}} \tag{5}$$

Neglecting the distinction between a_x and a_y, the generic form of the observation equations becomes:

$$\mathbf{F}\cdot a = b, \qquad \text{with: } \mathbf{F}^{M\times N}, a^{N\times 1}, b^{M\times 1} \tag{6}$$

Usually, (6) is an overdetermined system ($M > N$) which cannot be solved exactly.

In least-squares parameter estimation, that particular vector $\hat{a}$ is chosen which minimizes the sum of squares $v^{\mathrm{T}}v$ of the *residual vector* v:

$$v = b - \mathbf{F}\cdot\hat{a} \tag{7}$$

The solution of (5) then follows by forming the *normal equations*:

$$(\mathbf{F}^T\mathbf{F})\hat{a} = \mathbf{F}^T b \tag{8}$$

which is a consistent set of equations, allowing inversion of the normal datamatrix $\mathbf{F}^T\mathbf{F}$:

$$\hat{a} = \mathbf{F}_L^{-1} b, \qquad \mathbf{F}_L^{-1} = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T \tag{9}$$

where $\mathbf{F}_L^{-1}$ is the left-inverse (Rao and Mitra, 1971) of $\mathbf{F}$, also called maximum-rank pseudo-inverse (Peters and Wilkinson, 1970) or l-inverse $\mathbf{F}^l$ (Rauhala, 1974, 1976). It is tacitly assumed that no parameters are perfectly correlated, since this would render $\mathbf{F}^T\mathbf{F}$ singular, requiring the use of generalized or pseudo-inverses (Peters and Wilkinson, 1970; Rao and Mitra, 1971; Rauhala, 1974, 1976).

The *residual standard deviation* $\hat{\sigma}$ follows from the minimized sum of squares in (7), after normalizing with the degree of freedom d.f. $= M - N$:

$$\hat{\sigma}^2 = v^T v/(M - N) \tag{10}$$

and an estimator of the parameter covariance matrix is known to be:

$$\mathrm{cov}(\hat{a}) = \hat{\sigma}^2(\mathbf{F}^T\mathbf{F})^{-1}$$

Generalization to more than one column $\hat{a}$ and b, respectively, is straightforward. Thus, only one left-inverse $\mathbf{F}_L^{-1}$ is required in the solution of (5), and the respective standard deviations and covariance matrices follow by evaluation of $v_x^T v_x$ and $v_y^T v_y$.

This estimation procedure requires the inversion of the $N \times N$ normal datamatrix $\mathbf{F}^T\mathbf{F}$. Especially in the case of polynomial approximation:

$$f(x,y) = (1,x,y,x^2,xy,y^2,x^3,x^2y,\ldots)^T \tag{11}$$

this is a notoriously ill-conditioned procedure from a numerical point of view, due to rounding, overflow and underflow effects. Instability usually increases with the value of the condition number $c_{\mathbf{X}}$ of a square matrix $\mathbf{X}$, that is, the ratio of the largest and smallest absolute characteristic values.

In the following sections, two methods equivalent to (9) are discussed which both avoid explicit formation of the normal equations (8). By combining these methods, numerical stability improves even more. The first employs certain regular distributions of the independent variables used to evaluate b, in combination with approximation functions which are *decomposable* into functions of one independent variable only, i.e.:

$$\{f_i(x,y)\} = \{f_{xi}(x) \cdot f_{yi}(y)\}, \qquad i = 1,\ldots,N \tag{12}$$

It has the favourable property that inversion of $\mathbf{F}^T\mathbf{F}$ is replaced by inversion of two smaller matrices $\mathbf{X}^T\mathbf{X}$ and $\mathbf{Y}^T\mathbf{Y}$, with condition numbers whose product equals $c_{\mathbf{F}^T\mathbf{F}}$.

The second method avoids the formation of the normal equations (8) by *triangularizing* the original overdetermined system (6):

It may be realized, that the matrix notation (20—22) completely avoids the conventional "monolinear" left-inverse $\mathbf{F}_L^{-1}$ which implies large savings in memory and calculation time requirements, even with respect to the Kronecker product form $(\mathbf{X}_L^{-1} \otimes \mathbf{Y}_L^{-1})$ in (23).

Generalization of the factorization process described above to three or more dimensions $x,y,z,\ldots$ is a straightforward procedure; Rauhala (1972, 1974, 1976) has devised a new notation in order to circumvent the limitations of conventional "Cayleyan" matrix algebra which only operates on scalars, vectors and matrices.

Such multidimensional factorization procedures are closely related to the separation method used for the solution of partial differential equations by series expansion: Fourier analysis in thermodynamics is a classic example. For numerical solutions, additional savings in computation time are feasible by employing multidimensional Fast Fourier Transform techniques (Brenner, 1969), at the cost of even more restrictions due to the choice of harmonic functions and evenly spaced grid points with highly compound values for $m_x,m_y,m_z,\ldots$, usually integral powers of 2.

Since least-squares solution of a general $M \times N$ equation system requires approximately $\frac{1}{2}N^2(N+M)$ multiplications and additions (using Gauss-Jordan elimination), computational savings by employing bilinear least squares rather than conventional monolinear least squares are quite impressive. In the case of a consistent set of equations (where $\mathbf{F}_L^{-1}$ is replaced by the regular inverse $\mathbf{F}^{-1}$), we have $M = N$, and (9) would require $N^3 = n_x^3 \cdot n_y^3$ multiplications and additions. Conversely, (21) requires only $n_x^3 + n_y^3$ of such elementary arithmetic operations. Thus, for x—y symmetry, we have $m_x = m_y = m = n_x = n_y = n$, and an approximate "savings" factor $N^3/n^3 = n^3$ may be calculated. For $n = 10$, this would imply a reduction in the required number of arithmetic operations by a factor 1000, for $n^2 = 100$ parameters.

As shown in the Appendix, the characteristic values of $\mathbf{F}^T\mathbf{F}$ are located at the diagonal of the Kronecker product $\Lambda_X \otimes \Lambda_Y$, where Λ_X and Λ_Y are the diagonal similarity transformations of $\mathbf{X}^T\mathbf{X}$ and $\mathbf{Y}^T\mathbf{Y}$, containing their respective characteristic values on their main diagonals. Thus, the condition number of $\mathbf{F}^T\mathbf{F}$ is found to be

$$c_{\mathbf{F}^T\mathbf{F}} = c_{\mathbf{X}^T\mathbf{X}} \cdot c_{\mathbf{Y}^T\mathbf{Y}} \tag{26}$$

which for the case of x—y symmetry reduces to:

$$c_{\mathbf{F}^T\mathbf{F}} = c_{\mathbf{X}^T\mathbf{X}}^2$$

Consequently, calculation of $\hat{\mathbf{A}}$ via bilinear approximation (20) may be expected to be numerically superior to assessment of $\hat{a}$ via monolinear approximation (9). Kratky (1976) has demonstrated this numerical advantage using polynomial approximation functions. Employing a square, uniform calibration mesh with $m_x = m_y = m = 9$, he found substantial deterioration in the conventional, monolinear procedure with respect to the bilinear method for

as noted by Greville (1961) and Rao and Mitra (1971, ch 2, p. 11)

Thus, the calculation of an $N \times M = n_x n_y \times m_x m_y$ left inverse is replaced with the left-inverse calculation for two smaller matrices of orders $n_x \times m_x$, and $n_y \times m_y$, respectively, with consequent savings in calculation time

This feature is the central aspect in the *array-algebra* of Rauhala (1972, 1974, 1976) who, like Rao and Mitra (1971, pp. 11, 25) realized that an equation system of the form:

$$(\mathbf{X} \otimes \mathbf{Y})a = b \tag{19}$$

may be written in matrix form:

$$\mathbf{X}\mathbf{A}\mathbf{Y}^{\mathrm{T}} = \mathbf{B} \tag{20}$$

if each element b_{ij} in the $m_x \times m_y$ matrix B is taken equal to element $b_{i+(j-1)m_x}$ in (16), and by similarly defining:

$$\mathbf{A} = \{a_{ij}\}, \qquad a_{ij} = a_{i+(j-1)n_x}, \qquad i = 1,\ldots,n_x; \qquad j = 1,\ldots,n_y$$

In essence, A and B are obtained by segmenting a into columns for common approximation functions f_{y_j}, and b for common observations in y_l^*, $l = 1,\ldots,m_y$. In this fashion, the least-squares estimate $\hat{\mathbf{A}}$ and residual matrix V follow as

$$\hat{\mathbf{A}} = \mathbf{X}_{\mathrm{L}}^{-1}\mathbf{B}\mathbf{Y}_{\mathrm{L}}^{-\mathrm{T}} \tag{21}$$

and:

$$\mathbf{V} = \mathbf{B} - \mathbf{X}\hat{\mathbf{A}}\mathbf{Y}^{\mathrm{T}} \tag{22}$$

since (21), by analogy to (19) and (20) may be written as.

$$\hat{a} = (\mathbf{X}_{\mathrm{L}}^{-1} \otimes \mathbf{Y}_{\mathrm{L}}^{-1})b = (\mathbf{X} \otimes \mathbf{Y})_{\mathrm{L}}^{-1}b = \mathbf{F}_{\mathrm{L}}^{-1}b \tag{23}$$

By analogy to (10), the residual variance estimate $\hat{\sigma}^2$ now becomes:

$$\hat{\sigma}^2 = \mathrm{Trace}(\mathbf{V}^{\mathrm{T}}\mathbf{V})/(m_x m_y - n_x n_y) \tag{24}$$

Unfortunately, conventional matrix notation fails to describe the covariance "matrix" $\mathrm{cov}(\hat{\mathbf{A}})$ of a parameter matrix $\hat{\mathbf{A}}$, yet, the advantages of dimensionwise factorization may be exploited in the Kronecker form

$$\mathrm{cov}(\hat{a}) = \hat{\sigma}^2(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1} = \hat{\sigma}^2((\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \otimes (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1})$$

Thus, the covariance between elements $\hat{a}_{ik}$ and $\hat{a}_{jl}$ in $\hat{\mathbf{A}}$ is the product of element (i,j) in $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ and element (k,l) in $(\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}$ times the residual variance $\hat{\sigma}^2$.

Using the parameter matrix $\hat{\mathbf{A}}$, the approximation function $f(x,y)$ in (2) may now be written as

$$f(x,y) = f_x^{\mathrm{T}}(x)\hat{\mathbf{A}}f_y(y) \tag{25}$$

---

*One may note that (16) is the standard FORTRAN ordering of matrix elements b_{ij} in B

$$\begin{bmatrix} \mathbf{R} \\ \cdots \\ \mathbf{O} \end{bmatrix} \hat{a} = \begin{bmatrix} b' \\ \cdots \\ \nu' \end{bmatrix} \rightarrow \hat{a} = \mathbf{R}^{-1} b', \; \nu^{\mathrm{T}} \nu = \nu'^{\mathrm{T}} \nu' \tag{13}$$

where inversion of R is reduced to a simple process of backsubstitution since R is upper triangular. This second method does not require special limitations on the class of approximation functions and calibration distribution. In addition, the condition number of R, at least for polynomial approximations, tends to be much smaller than $(c_{\mathbf{F}^{\mathrm{T}}\mathbf{F}})^{1/2}$, for moderate and high values of M and N.

BILINEAR LEAST SQUARES

If the approximation functions $f_i(x,y)$ are decomposable as in (12), (1) may be ordered as the *Kronecker product* (Greville, 1961; Rao and Mitra, 1971, p. 11) of the individual approximation vectors in x and y:

$$\begin{aligned} f(x,y) &= (f_x^{\mathrm{T}}(x) \cdot f_{y1}(y), \ldots, f_x^{\mathrm{T}}(x) \cdot f_{yn_y}(y))^{\mathrm{T}} \\ &\equiv f_x(x) \otimes f_y(y) \end{aligned} \tag{14}$$

with:

$$f_x(x) = (f_{x1}(x), \ldots, f_{xn_x}(x))^{\mathrm{T}}$$

$$f_y(y) = (f_{y1}(y), \ldots, f_{yn_y}(y))^{\mathrm{T}}$$

For a discussion of this rather unknown matrix product, the reader is referred to the Appendix. In essence, each element of the second factor is replaced by its product with the first factor.

Let observation matrices X and Y be defined as:

$$\begin{aligned} \mathbf{X} &= [f_x(x_1) \vdots \cdots \vdots f_x(x_M)]^{\mathrm{T}} \\ \mathbf{Y} &= [f_y(y_1) \vdots \cdots \vdots f_y(y_M)]^{\mathrm{T}} \end{aligned} \tag{15}$$

and let the observations be taken for independent variables contained at the junctions of a rectangular grid:

$$\{(x_i, y_j)\}; \qquad i = 1, \ldots, m_x; \qquad j = 1, \ldots, m_y$$

If the observations are ordered with increasing i before j, i.e.:

$$b = ((b_{i+(j-1)m_x}, i = 1, \ldots, m_x), j = 1, \ldots, m_y)^{\mathrm{T}} \tag{16}$$

F in (5) is found to reduce to the Kronecker product:

$$\mathbf{F} = \mathbf{X} \otimes \mathbf{Y} \tag{17}$$

where again each element of the second factor is replaced by its product with the first factor.

The l- or left-inverse $\mathbf{F}_{\mathrm{L}}^{-1}$ in (9) then follows as:

$$\mathbf{F}_{\mathrm{L}}^{-1} = \mathbf{X}_{\mathrm{L}}^{-1} \otimes \mathbf{Y}_{\mathrm{L}}^{-1} \tag{18}$$

$n_x = n_y = n > 5$, using single precision (24 bits mantissa) floating point arithmetics in a PDP-11/45 computer*. In addition, time savings for the latter procedure were considerable, such as a factor 12.6 for $n = 5$.

In conventional, monolinear least squares the influence of blunders in observations is countered by removing observations whose absolute residual is larger than, say, $3\hat{\sigma}$, and by repeating the calibration procedure until no further removals are necessary. In bilinear least squares, one would have to remove a complete row or column of observations. Alternatively, one might replace an offending observation with its approximated value and repeat the calibration procedure until no further improvement is obtained.

Bilinear least squares may be generalized to accommodate weight matrices. Again, the structured form of the Kronecker product $\mathbf{X} \otimes \mathbf{Y}$ implies that weight matrices will affect complete rows and columns, rather than points individually. Similarly, entering a new approximation component $f_{xn_x+1}(x)$, say, will entail the introduction of a complete *vector* $f_{xn_x+1}(x) \cdot f_y(y)$ into the observation equations (20).

Apart from the restrictions on this structure, one is free to choose approximation functions and calibration grid lay-out. In particular, grid spacing does not have to be uniform. Kratky (1976) points out, that non-uniform meshes may be more efficient in those frequent cases where image errors and their spatial derivatives increase towards the image edges or corners. Decreasing mesh spacings toward the image perifery may allow acceptable sampling dimensions m_x and m_y which are smaller than would be required for uniform grids.

ORTHOGONAL TRIANGULARIZATION

Another method avoiding explicit formation of the conventional, normal observation matrix $\mathbf{F}^T\mathbf{F}$ is based on *orthogonal*** transformation of the original datamatrix F. By premultiplying F with a suitably chosen orthonormal matrix Q, i.e.:

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I} \tag{27}$$

the resultant product QF may adopt upper triangular form:

*The condition numbers were, respectively:

$$c_{\mathbf{X}^T\mathbf{X}} = c_{\mathbf{Y}^T\mathbf{Y}} = \begin{cases} 9.9 \cdot 10^5 \\ 3.0 \cdot 10^7 \end{cases}; \qquad c_{\mathbf{F}^T\mathbf{F}} = \begin{cases} 9.9 \cdot 10^{11} & (n = 6) \\ 9.3 \cdot 10^{14} & (n = 7) \end{cases}$$

**Strictly speaking, an *orthogonal* matrix Q has columns q_i which are mutually orthogonal ($q_i^T q_j = 0$ for $i \neq j$) implying that $\mathbf{Q}^T\mathbf{Q}$ is diagonal. In an *orthonormal* matrix, these columns have also unit length, i.e. $q_i^T q_j = \delta_{ij}$, which implies (27). Unfortunately, the term orthogonal is often used where orthonormality is meant, such as in the expression "orthogonal triangularization". Unless confusion may arise, the conventional terms are adopted.

$$\mathbf{QF} = \begin{bmatrix} \mathbf{R} \\ \cdots \\ \mathbf{O} \end{bmatrix}, \text{ with } \mathbf{R} = \begin{bmatrix} r_{11} & \cdots & r_{1N} \\ & \ddots & \vdots \\ 0 & & r_{NN} \end{bmatrix} \tag{28}$$

By applying $\mathbf{Q}$ to both sides of (6), one finds:

$$(\mathbf{QF})a = \mathbf{Q}b \tag{29}$$

This transformation has no influence on the least-squares estimate $\hat{a}$, as one may show by explicit formation of the normal equations:

$$(\mathbf{QF})^{\mathrm{T}}(\mathbf{QF})\hat{a} = (\mathbf{QF})^{\mathrm{T}}\mathbf{Q}b$$

which reduces to (8) by virtue of the orthonormality property (27). Consequently, (29) may be written, using (28), as:

$$\begin{bmatrix} \mathbf{R} \cdot \hat{a} \\ \cdots \\ \mathbf{O} \end{bmatrix} = \mathbf{Q}b = \begin{bmatrix} b' \\ \cdots \\ v' \end{bmatrix} \tag{30}$$

which shows that the upper N rows form a consistent system of equations, and where the lower $(M - N)$ elements of $\mathbf{Q}b$ contain the nonzero elements of the transformed residual vector v'. In a similar fashion as for $\hat{a}$, one may prove that the residual sum of squares is not affected by this orthogonal transformation. Consequently, the least-squares solution $\hat{a}$ and residual variance $\hat{\sigma}^2$ follow as:

$$\hat{a} = \mathbf{R}^{-1}b', \qquad \hat{\sigma}^2 = v'^{\mathrm{T}}v'/(M - N) \tag{13}$$

Since $\mathbf{R}$ is upper triangular, solution of $\mathbf{R}^{-1}b'$ is a simple process of backsubstitution, i.e. first $\hat{a}_N$ is solved, subsequently $\hat{a}_{N-1}$ after substitution of $\hat{a}_N$ into the first $N - 1$ equations, etc., until $\hat{a}_1$ is finally obtained.

Yassa (1974) has discussed the application of orthogonal transformations in analytical photogrammetry; he demonstrated that $\mathbf{R}$ is the Choleski decomposition of $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ which is often used in symmetrical matrix inversion as required for $\mathbf{F}_{\mathrm{L}}^{-1}$ in (9). Thus, rather than forming normal datamatrices which are subsequently decomposed into triangular factors for inversion by backsubstitution, orthonormal transformations allow direct assessment of these triangular matrices. Since the condition number of an orthonormal matrix is equal to 1 (the length of a vector being invariant with an orthonormal transformation), the transformation (29) may be expected to be numerically stable.

A relation such as (26) between the condition numbers of $\mathbf{R}$ and $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ $(=\mathbf{R}^{\mathrm{T}}\mathbf{R})$ cannot be given for arbitrary $\mathbf{R}$. Only for the trivial case of diagonal $\mathbf{R}$, a squaring relation obtains since then $\mathbf{R} = \mathbf{R}^{\mathrm{T}}$, and since squaring of a matrix entails squaring of its characteristic values. However, numerical evidence below suggests that $c_{\mathbf{R}}^2 << c_{\mathbf{F}^{\mathrm{T}}\mathbf{F}}$ in the case of polynomial datamatrices of type (15) and (17), for moderate and high polynomial orders.

A number of procedures for explicitly obtaining $\mathbf{Q}$ or for directly generating $\mathbf{R}$ and b' are described in the literature, such as Householder transformations

(Householder, 1958; Peters and Wilkinson, 1970), Modified Gram-Schmidt orthonormalization (Peters and Wilkinson, 1970) and Givens rotations (Givens, 1958). Yassa (1974) has discussed the former two. Whereas these former methods are typically applied to the columns of **F** and b, Givens rotations may be applied in a *rowwise* fashion, which allows stepwise row acquisition in highly overdetermined equation systems, thus reducing memory requirements. In addition, sparse matrix structures are readily exploited with beneficial effects on calculation time (Bunch and Rose, 1976).

Gentleman (1973) has described a particularly attractive procedure based on Givens rotations, in which the calculation of square roots is avoided. Basically, **F** is transformed in such a way as to obtain:

$$\mathbf{QF} = \begin{bmatrix} \mathbf{D}^{1/2}\bar{\mathbf{R}} \\ \cdots \\ \mathbf{O} \end{bmatrix}$$

with **D** diagonal, and $\bar{\mathbf{R}}$ upper unit triangular, i.e. containing ones in its main diagonal, implying that $c_{\bar{\mathbf{R}}} = 1$. His algorithm generates **D** and $\bar{\mathbf{R}}$, rather than $\mathbf{R} = \mathbf{D}^{1/2}\bar{\mathbf{R}}$. By treating b in (29) as an additional column appended to **F**, the transformed column b' of (30) is obtained in the factorized form $\mathbf{D}^{1/2}\bar{b}'$, together with the residual sum of squares $v^{\mathrm{T}}v = v'^{\mathrm{T}}v'$ as an additional element of **D**.

Thus, $\hat{a}$ and $\hat{\sigma}^2$ are obtained as:

$$\hat{a} = \bar{\mathbf{R}}^{-1}\bar{b}', \qquad \hat{\sigma}^2 = v^{\mathrm{T}}v/(M - N)$$

Peters and Wilkinson (1970) have suggested that ill-conditioning of **F** is usually absorbed in **D**. Therefore, additional numerical stability may be expected, which is borne out by the unit value of $c_{\bar{\mathbf{R}}}$. In addition, $\hat{\sigma}^2$ is an automatic by-product of this calculation procedure, although ill-conditioning of **D** may require some caution in the use of this estimate.

BILINEAR LEAST SQUARES BY ORTHOGONAL TRIANGULARIZATION

Generalization of such triangularization procedures to more than one column b, i.e. a matrix, is trivial; for each appended observation column, a corresponding parameter column may be obtained. This allows an elegant application of orthogonal triangularization to the bilinear form (20):

$$\mathbf{Q_X}(\mathbf{XAY}^{\mathrm{T}})\mathbf{Q_Y}^{\mathrm{T}} = (\mathbf{Q_X X})\mathbf{A}(\mathbf{Q_Y Y})^{\mathrm{T}} = \mathbf{Q_X B Q_Y}^{\mathrm{T}} \tag{31}$$

leaving $\hat{\mathbf{A}}$ unmodified since the corresponding monolinear transformation is:

$$(\mathbf{Q_X} \otimes \mathbf{Q_Y})\mathbf{F}a = (\mathbf{Q_X} \otimes \mathbf{Q_Y})b$$

which may be shown to be orthonormal by inspection.

Despite the numerical accuracy of these methods, traditional tools such as scaling and centering of independent variables retain their importance. For the case of uniform grids, i.e.:

$x_i - x_{i-1} = \Delta x, \qquad y_j - y_{j-1} = \Delta y$ (constant)

Rauhala (1972) suggested to scale the independent variables in terms of these increments, and to center to their mean value. Thus, for odd m, the independent variables take the values $0, \pm 1, \pm 2, \ldots$, and for even m, the values $\pm\frac{1}{2}, \pm\frac{3}{2}, \ldots$.

Using polynomial approximations, the datamatrices X and Y then have the following forms:

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -2 & 4 & -8 & 16 & . & . \\ 1 & -1 & 1 & -1 & 1 & . & . \\ 1 & 0 & 0 & 0 & 0 & . & . \\ 1 & 1 & 1 & 1 & 1 & . & . \\ 1 & 2 & 4 & 8 & 16 & . & . \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \qquad \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -\frac{3}{2} & \frac{9}{4} & -\frac{27}{8} & \frac{81}{16} & . & . \\ 1 & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{8} & \frac{1}{16} & . & . \\ 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & . & . \\ 1 & \frac{3}{2} & \frac{9}{4} & \frac{27}{8} & \frac{81}{16} & . & . \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} +$$

odd m even m

For large equation systems, X and Y may be rendered sparse by the following orthogonal transformations (the condition numbers c_R are not affected by this operation, since triangular factorizations R are unique for some given matrix $\mathbf{F}^T\mathbf{F}$, and since the latter is invariant to an orthogonal transformation of F):

$$\begin{bmatrix} . & . & & \vdots & & . & . \\ & & -\frac{1}{2}\sqrt{2} & 0 & \frac{1}{2}\sqrt{2} & & \\ & . & 0 & 1 & 0 & . & . \\ & & \frac{1}{2}\sqrt{2} & 0 & \frac{1}{2}\sqrt{2} & & \\ . & . & & \vdots & & . & . \end{bmatrix} \qquad \begin{bmatrix} . & . & & \vdots & \vdots & & . \\ & & -\frac{1}{2}\sqrt{2} & 0 & 0 & \frac{1}{2}\sqrt{2} & \\ & . & 0 & -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 & . \\ & . & 0 & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 & . \\ & & \frac{1}{2}\sqrt{2} & 0 & 0 & \frac{1}{2}\sqrt{2} & \\ . & . & & \vdots & \vdots & & . \end{bmatrix}$$

odd m even m

which results in the transformed matrices:

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 2 & 0 & 8 & 0 & . & . \\ 0 & 1 & 0 & 1 & 0 & . & . \\ 1 & 0 & 0 & 0 & 0 & . & . \\ 1 & 0 & 1 & 0 & 1 & . & . \\ 1 & 0 & 4 & 0 & 16 & . & . \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \qquad \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \frac{3}{2} & 0 & \frac{27}{8} & 0 & . & . \\ 0 & \frac{1}{2} & 0 & \frac{1}{8} & 0 & . & . \\ 1 & 0 & \frac{1}{4} & 0 & \frac{1}{16} & . & . \\ 1 & 0 & \frac{9}{4} & 0 & \frac{81}{16} & . & . \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

odd m even m

The original $m \times n$ least-squares problem (per dimension) is replaced by two uncorrelated $\frac{1}{2}m \times \frac{1}{2}n$ problems, which may be either solved in a completely

separated sense or by exploiting skipping rules in subsequent triangularization. In the following section, some timing results of the last approach are presented.

Of course, for routine use it may be worth while to save the left-inverses $\mathbf{X}_L^{-1}$ and $\mathbf{Y}_L^{-1}$, for direct evaluation of $\hat{\mathbf{A}}$ in (21), as Kratky (1976) has suggested. Using the triangularization approach, $\mathbf{X}_L^{-1}$ follows as:

$$\mathbf{X}_L^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = (\mathbf{R}^T\mathbf{R})^{-1}[\mathbf{R}^T:\mathbf{O}]\mathbf{Q} = [\mathbf{R}^{-1}:\mathbf{O}]\mathbf{Q}$$

where **R** and **Q** are obtained in the triangularization procedure:

$$\mathbf{XA} = \mathbf{I} \rightarrow \begin{bmatrix} \mathbf{R} \\ \cdots \\ \mathbf{O} \end{bmatrix} \mathbf{A} = \mathbf{Q}$$

EXPERIMENTAL RESULTS

The theory discussed in the previous sections has been verified in calibration applied to image distortion in a SELSPOT (Anonymous, 1975) system for real-time movement monitoring. This optoelectronic camera system observes the direction of a number of infrared light emitting diodes (IR LEDs) which are operated in time division multiplex. A continuously position sensitive photodiode (Woltring, 1975a) converts the mean position of an incident light distribution into photocurrents which are amplified and normalized in order to compensate for signal light intensity variations. Background light influence is cancelled by operating on output current changes when an LED is turned on. Output data x and y are available in digital 10-bits words (range −511 to +511), which implies a formal quantization noise, under the assumption of a uniform noise distribution, with $\sigma = (1/12)^{1/2} \cong 0.3$ LSB (Least Significant Bit).

In previous work (Woltring, 1975b), using standard, non-metric CCTV optics (CANON TV-16 fixed-focus lens 50 mm/0.95—2.2), radial distortion was found to range between 3% and 6% full scale in the image corners, for aperture settings ranging between 0.95 and 2.2. In the present experiment, raw data were obtained by averaging 314 observation on an LED (i.e., 1-sec measurement interval), in order to reduce the influence of quantization noise and photodetector noise. The LED was manually moved into the holes of a hexagonally perforated metal sheet with dimensions 2 m × 1 m. A rectangular grid pattern was chosen with horizontal grid distance 30 mm, and vertical grid distance $30\sqrt{3/4}$ mm. A SELSPOT camera was placed at approximately 2.2 m distance from the calibration plane, with its image axes oriented approximately in parallel with those of the calibration grid. Thus, some 70% image coverage was attained in the vertical direction.

Using a 32 × 37 grid, an approximately square image was obtained. The camera aperture was set at 1.4, and the distance at infinity (IR mark), thus causing some image blurring which has no influence on system resolution because of the integrating properties of the position sensitive photodetector. In this fashion, the dominating pincushion distortion exhibited by the CANON

lens is slightly offset, and any local irregularities in the position sensitive detector are smoothed. Data were fed into a PDP-11/45 minicomputer for subsequent bilinear least-squares processing.

In calibration, actual calculation times for the parameter matrices $\hat{\mathbf{A}}_x$ and $\hat{\mathbf{A}}_y$ (jointly assessed by analogy to (5)) and residual standard deviation estimates $\hat{\sigma}_x$ and $\hat{\sigma}_y$ have been calculated; the latter were obtained by explicit evaluation of the residual matrices $\mathbf{V}_x$ and $\mathbf{V}_y$ (22). The following procedures have been compared: use of normal equations per dimension (18), and triangularization using Gentleman's version of Givens rotations with and without presparsening of the datamatrices $\mathbf{X}$ and $\mathbf{Y}$.

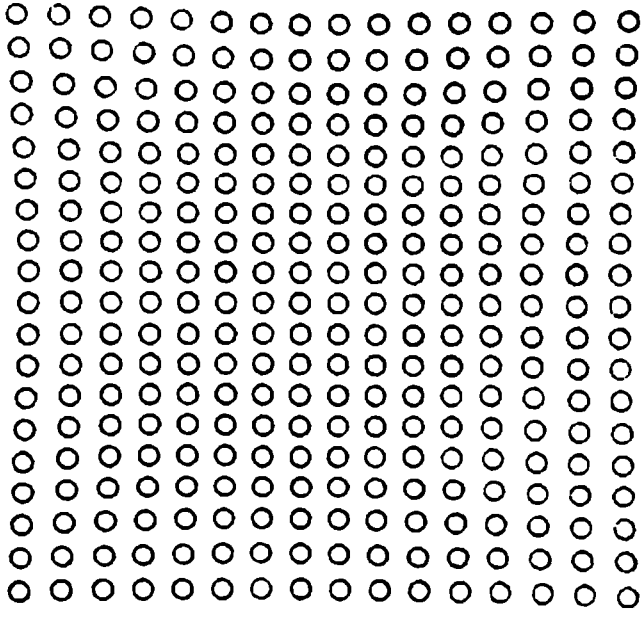


Fig. 1. Distorted image for low-density calibration grid.

Two cases have been selected as an example: low and high mesh density, where the former is obtained by rejection of every other row and column in the high-density grid (see Fig. 1): Polynomial terms up to the 9th power have been used, entailing up to 100 parameters. For all procedures, single precision floating point arithmetics was employed, except for: (1) inner vector products were accumulated in double precision, and (2) inversion of the normal datamatrices $\mathbf{X}^T\mathbf{X}$ and $\mathbf{Y}^T\mathbf{Y}$ took place via Choleski decomposition (no pivoting),

with the Choleski factors calculated in double precision. Since these comparison experiments were run on a PDP-11/45 without floating point hardware (RT-11 F/B operating system), the influence of floating point multiplications on calculation times is considerable.

It was found that the residual standard deviations as assessed by the two triangularization procedures yielded virtually identical results, with incidental unit differences in the 5th decimal; therefore, Tables I and II distinguish only the residual standard deviations for normal equations and for triangularization without presparsening. The former procedure exhibits increasing numerical deterioration for $n > 6$, as is borne out by the large ratios of the condition numbers of the respective normal and triangular datamatrices. Since actual calculations in triangularization employed Gentleman's $D^{1/2}\bar{R}$ decomposition

TABLE I

High-density calibration grid

$m_x = 32$ $m_y = 37$		Condition numbers		Residual standard deviations		Parameter calculation times (excl. s.d.)		
A	B	C	D	C	D	C	D	E
0	1	1 1	1 1	236 LSB 221	236 LSB 221	3.1 s	5.0 s	5.9 s
1	4	8.5×10^{1} 1.1×10^{2}	9.2×10^{0} 1.1×10^{1}	4.99 4.56	4.99 4.56	6.5	8.3	7.7
2	9	2.9×10^{4} 5.3×10^{4}	7.6×10^{1} 1.0×10^{2}	3.61 3.61	3.61 3.61	10.4	11.7	9.5
3	16	5.4×10^{6} 1.3×10^{7}	6.2×10^{2} 9.5×10^{2}	1.05 1.18	1.05 1.18	14.8	15.3	11.3
4	25	1.7×10^{9} 3.8×10^{9}	4.9×10^{3} 8.8×10^{3}	0.92 1.11	0.92 1.11	19.8	19.1	13.3
5	36	3.4×10^{11} 1.5×10^{12}	3.9×10^{4} 8.1×10^{4}	0.79 0.77	0.79 0.77	25.3	23.0	15.4
6	49	6.8×10^{13} 4.1×10^{13}	3.1×10^{5} 7.4×10^{5}	0.65300 0.66674	0.65299 0.66675	31.4	27.1	17.6
7	64	1.1×10^{16} 8.2×10^{16}	2.4×10^{6} 6.8×10^{6}	0.53766 0.44714	0.53700 0.44701	38.4	31.5	19.9
8	81	? ?	1.9×10^{7} 6.1×10^{7}	0.47245 0.42697	0.46624 0.42687	45.9	36.0	22.3
9	100	? ?	1.4×10^{8} 5.5×10^{8}	1.0968 0.9502	0.43666 0.36290	56.7	40.7	24.8

A = polynomial order; *B* = number of parameters; *C* = normal; *D* = triangularization; *E* = triangularization with presparsening.
s.d. = residual standard deviation.

TABLE II

Low-density calibration grid

$m_x = 16$ $m_y = 19$		Condition numbers		Residual standard deviations		Parameter calculation times (excl. *s.d.*)		
A	*B*	*C*	*D*	*C*	*D*	*C*	*D*	*E*
0	1	1 1	1 1	236 LSB 227	236 LSB 227	0.9 s	1.3 s	1.6 s
1	4	2.1×10^{1} 3.0×10^{1}	4.6×10^{0} 5.5×10^{0}	5.20 4.69	5.20 4.69	2.0	2.3	2.1
2	9	1.8×10^{3} 3.7×10^{3}	1.9×10^{1} 2.7×10^{1}	3.44 3.82	3.44 3.82	3.3	3.3	2.6
3	16	8.3×10^{4} 2.3×10^{5}	7.5×10^{1} 1.3×10^{2}	1.07 1.20	1.07 1.20	5.0	4.4	3.2
4	25	6.4×10^{6} 2.5×10^{7}	2.9×10^{2} 5.9×10^{2}	0.94 1.13	0.94 1.13	6.9	5.6	3.9
5	36	3.3×10^{8} 1.8×10^{9}	1.1×10^{3} 2.7×10^{3}	0.86980 0.74548	0.86981 0.74548	9.1	6.9	4.6
6	49	2.9×10^{10} 1.7×10^{11}	4.2×10^{3} 1.2×10^{4}	0.65400 0.65340	0.65400 0.65339	11.7	8.4	5.4
7	64	1.3×10^{12} 1.6×10^{13}	1.5×10^{4} 5.5×10^{4}	0.49532 0.37584	0.49267 0.37517	14.6	9.9	6.3
8	81	4.0×10^{13} 5.4×10^{14}	5.2×10^{4} 2.4×10^{5}	0.39600 0.38153	0.39053 0.36043	18.0	11.6	7.3
9	100	3.2×10^{15} 4.3×10^{16}	1.7×10^{5} 9.9×10^{5}	3.7764 0.3742	0.37231 0.34126	21.6	13.3	8.3

A = polynomial order; *B* = number of parameters; *C* = normal; *D* = triangularization; *E* = triangularization with presparsening.
s.d. = residual standard deviation.

for R_X and R_Y, this ratio is in fact even more pronounced — assuming complete absorption of ill-conditioning of X and Y into D_X and D_Y, respectively.

Only for $n = 1$ and 2, the normal equations approach is faster than the two other ones. For large n, the triangularization procedure with presparsening is more than twice as fast as the normal equations method, with the other triangularization procedure ranging between them.

In these experiments, the averaged raw data were directly entered into the bilinear approximation routines. One may expect that additional numerical improvements are feasible if a trend function is fitted first, e.g. by two subsequent bilinear approximations. Thus, one might first assess a residual matrix V for $n_x = n_y = 2$, and use this as input B for a second approximation with higher polynomial orders.

DISCUSSION

Results such as those in the previous section suggest that the combination of bilinear approximation and orthogonal triangularization is most suitable in the case of high dimensions n_x and n_y. Referring to Rauhala's (1972, 1974, 1976) and Kratky's (1976) findings, both calculation speed and numerical accuracy are improved as compared to bilinear least squares using normal equations. In addition, neither normal equations nor orthogonal triangularization in monolinear least squares are suited for very high numbers of observations m_x and m_y.

A disadvantage of calibration using rectangular grids is that usually the wrong variables are available in this regular form. In photogrammetry, one typically wishes to use distorted image points as independent variables in order to recover the original, undistorted image as a set of dependent variables. However, when using reseau cameras or calibration grids in object space, the image will exhibit a distorted grid.

One approach to cope with this problem is to define a rectangular grid on the distorted image, and to interpolate by conventional means on say the 25 closest observations, as obtained via arbitrary calibration distributions, using image coordinates as independent variables. This may be comparatively efficient, since for each new mesh point — when processed in a suitable order — most of the observation equations used for the previous mesh point remain valid. In fact, Gentleman (1973) discussed a modification of Givens rotations which allows removal of previously processed observations.

Alternatively, one may assess approximation parameters using the wrong order of dependent and independent parameters, and calculate iteratively a rectangular mesh in the distorted image. On this grid, bilinear approximation may again be applied.

Yet, there are instances where distorted image coordinates may serve as dependent variables. In our situation, the use of Extended Kalman filtering for spatial reconstruction of dynamic processes constitutes one example (Woltring, 1976). Since new observations on moving targets are processed after linearization of observation equations around current estimates of object space coordinates, the use of calibration grids in object space becomes appropriate. However, Extended Kalman filtering is a very time-consuming process, and various suboptimal filtering methods may require image distortion calibration in the reverse form.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Vladimir Kratky of NRC Canada and Dr. Urho A. Rauhala of DBA Systems, Florida, for stimulating discussions on analytical methods in Photogrammetry, during a recent visit to Canada and the United States of America which was made possible in part through a travel

grant from the Netherlands Organisation for the Advancement of Pure Research (Z.W.O.). In addition, he acknowledges their criticisms on earlier drafts of this paper. Finally, he is indebted to Ir. A. Van den Boogaard and Mr. C.M. Grootendorst of ITC Enschede, who made available to him the software employed in this research for condition number calculation of symmetrical matrices.

APPENDIX

The *Kronecker product* of two arbitrary matrices $\mathbf{X} = [x_{ij}]$, $i = 1,\ldots,m_x$, $j = 1,\ldots,n_x$, and $\mathbf{Y} = [y_{ij}]$, $i = 1,\ldots,m_y$, $j = 1,\ldots,n_y$ is defined as:

$$\mathbf{X} \otimes \mathbf{Y} \equiv [\mathbf{X}y_{ij}] = \begin{bmatrix} \mathbf{X}y_{11} & \cdots & \mathbf{X}y_{1n_y} \\ \vdots & & \vdots \\ \mathbf{X}y_{m_y1} & \cdots & \mathbf{X}y_{m_yn_y} \end{bmatrix} \tag{A1}$$

an $m_x m_y \times n_x n_y$ matrix expressible as a partitioned matrix with submatrix $\mathbf{X}y_{ij}$ in its (i,j)th partition (Rao and Mitra, 1971)*. Smirnov (1964) calls it the "direct product of matrices", denoted as $\mathbf{X} \times \mathbf{Y}$ (Rauhala, 1974).

A variety of operations on a Kronecker product may be replaced by operations on the constituent factors, with large savings in computational requirements as regards memory space and calculation time. The following properties are stated without proof (Rao and Mitra, 1971); they may easily be verified by inspection, if the product (A1) is explicitly written out.

Identity: $\mathbf{I} \otimes \mathbf{I} = \mathbf{I}$ (A2)

Zeroing: $\mathbf{A} \otimes \mathbf{O} = \mathbf{O} \otimes \mathbf{A} = \mathbf{O}$ (A3)

Transposition: $(\mathbf{X} \otimes \mathbf{Y})^{\mathrm{T}} = \mathbf{X}^{\mathrm{T}} \otimes \mathbf{Y}^{\mathrm{T}}$ (A4)

Distributivity: $(\mathbf{X}_1 + \mathbf{X}_2) \otimes (\mathbf{Y}_1 + \mathbf{Y}_2) = \mathbf{X}_1 \otimes \mathbf{Y}_1 + \mathbf{X}_2 \otimes \mathbf{Y}_1 + \mathbf{X}_1 \otimes \mathbf{Y}_2 + \mathbf{X}_2 \otimes \mathbf{Y}_2$ (A5)

Associativity: $(a\mathbf{X}) \otimes (b\mathbf{Y}) = (ab)(\mathbf{X} \otimes \mathbf{Y})$ (A6)

$(\mathbf{X}_1\mathbf{X}_2) \otimes (\mathbf{Y}_1\mathbf{Y}_2) = (\mathbf{X}_1 \otimes \mathbf{Y}_1)(\mathbf{X}_2 \otimes \mathbf{Y}_2)$ (A7)

Through these relations, the following are easily proven:

Regular inverse: $(\mathbf{X} \otimes \mathbf{Y})^{-1} = \mathbf{X}^{-1} \otimes \mathbf{Y}^{-1}$ (if they exist) (A8)

Left inverse: $(\mathbf{X} \otimes \mathbf{Y})_{\mathrm{L}}^{-1} = \mathbf{X}_{\mathrm{L}}^{-1} \otimes \mathbf{Y}_{\mathrm{L}}^{-1}$ (if they exist) (A9)

Normal matrix: If $\mathbf{F} = \mathbf{X} \otimes \mathbf{Y}$, $\mathbf{F}^{\mathrm{T}}\mathbf{F} = (\mathbf{X}^{\mathrm{T}}\mathbf{X}) \otimes (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})$ (A10)

---

*Rauhala pointed out in a personal communication (1976), that Rao and Mitra (1971, p. 11) erroneously define $\mathbf{X} \otimes \mathbf{Y}$ as $[x_{ij}\mathbf{Y}]$. This would entail rowwise rather than columnwise ordering of the corresponding monolinear vectors in (19) and (20), whereas Rao and Mitra (1971, p. 25) talk about columnwise expansion.

Characteristic values: if Q_x and Q_y are similarity transformations rendering X^TX and Y^TY diagonal,

$$\Lambda_X = Q_x(X^TX)Q_X^{-1}, \qquad \Lambda_Y = Q_Y(Y^TY)Q_Y^{-1}$$

the Kronecker product $\Lambda_X \otimes \Lambda_Y$ is the diagonal similarity transformation of $F^TF = (X^TX) \otimes (Y^TY)$, since:

$$\begin{aligned}\Lambda_X \otimes \Lambda_Y &= (Q_X(X^TX)Q_X^{-1}) \otimes (Q_Y(Y^TY)Q_Y^{-1}) \\ &= (Q_X \otimes Q_Y)((X^TX) \otimes (Y^TY))(Q_X \otimes Q_Y)^{-1}\end{aligned}$$

as may be shown by repeated application of (A7).

As a consequence, this Kronecker product contains the characteristic values of F^TF on its main diagonal, since Λ_X and Λ_Y contain the characteristic values of the constituent factors X^TX and Y^TY on their main diagonals.

REFERENCES

Anonymous, 1975. Photodetector picks up LED emissions to measure positions of moving objects. Electronics, 48(4): 6E—8E.

Brenner, N.H., 1969. Fast Fourier transform of externally stored data. IEEE Trans. Audio Electroacoust., 17: 128—132.

Bunch, J.R. and Rose, D.J. (Editors), 1976. Sparse Matrix Computations. Academic Press, New York, N.Y.

Gentleman, W.M., 1973. Least squares computations by Givens transformations without square roots. J. Inst. Math. Appl., 12: 329—336.

Greville, T.N., 1961. Note on fitting of functions of several independent variables. J. Soc. Ind. Appl. Math., 9: 109—115.

Givens, J.W., 1958. Computation of plane unitary rotations transforming a general matrix to triangular form. J. Soc. Ind. Appl. Math., 6: 26—50.

Householder, A.S., 1958. Unitary triangularization of a nonsymmetrical matrix. J. Assoc. Comput. Mach., 5: 339—342.

Kratky, V., 1976. Grid-modified polynomial transformation of satellite imagery. Remote Sensing Environ., 5: 67—74.

Peters, G. and Wilkinson, J.H., 1970. The least-squares problem and pseudo-inverses. Comput. J., 13: 309—316.

Rao, C.R. and Mitra, S.K., 1971. Generalized Inverse of Matrices and its Applications. Wiley, New York, N.Y.

Rauhala, U.A., 1972. Calculus of Matrix Arrays and General Polynomial and Harmonic Interpolation by Least Squares with New Solutions in Photogrammetry and Geodesy. Photogramm. Medd., 6(4). (Thesis, Royal Institute of Technology, Stockholm.)

Rauhala, U.A., 1974. Array Algebra with Applications in Photogrammetry and Geodesy. Photogramm. Medd., 6(6) (Thesis, Royal Institute of Technology, Stockholm).

Rauhala, U.A., 1976. A review of array algebra. Photogramm. Medd., 2: 38.

Smirnov, V.I., 1964. A Course of Higher Mathematics, III. Addison-Wesley, London.

Woltring, H.J., 1975a. Single- and dual-axis lateral photodetectors of rectangular shape. IEEE Trans. Electron Devices, 22: 581—590, 1101.

Woltring, H.J., 1975b. Calibration and measurement in 3-dimensional monitoring of human motion by optoelectronic means, I. Preliminaries and theoretical aspects. Biotelemetry, 2: 169—196.

Woltring, H.J., 1976. Calibration and measurement in 3-dimensional monitoring of human motion by optoelectronic means, II. Experimental results and discussion. Biotelemetry, 3(2): in press.

Yassa, G., 1974. Orthogonal transformations in triangulation adjustment. Photogramm. Eng., 40: 961—966.

Chapter III

Experiments in Human Motor Control

If I were running in the stadium, ought I to slacken my pace when approaching the goal? Ought I not rather to put on speed?

Diogenes, when told that he should take a rest since he was an old man.

3.1

Modifiability of a 1-Dimensional, Fast Step Response *

Summary

A common distinction in the study of fast arm movements to a target is between an initial, "ballistic" or open-loop, distance-covering phase, and a subsequent, closed-loop or "homing"-phase to the target However, little is known about which part of all incoming information during movement execution is or may be used for this latter control phase

In this paper, a number of experiments are reported in which feedback mode, gain and availability are experimentally controlled. Auditory and visual feedback were employed (pitch, visual light spot distance on a display), feedback gain could be varied within or between series of step tracking trials, and feedback could be made available during specific movement phases only Subjects had to cover distances ranging between 20 and 50 cm, they were prompted to move as fast as possible and yet to attain high targeting accuracy. Data were analyzed through linear regression models, the slope of the regression line relating required and actual distances appeared to be quite sensitive to various experimental treatments.

It was found that visual feedback was more beneficial than auditory feedback, despite the shorter reaction times to the latter type of feedback, for movement times ranging between 500 and 800 msec (on average). For gradual target information presentation over short initial movement intervals, the results suggested the adoption of an information accumulating strategy minimizing central capacity processing requirements, since integration of information common to all trials within a series was found For larger initial feedback intervals, differentiation in target information processing was observed. Finally, for sudden target changes during initial movement phases, a differential effect was found between larger and shorter distances to be covered, suggesting the adoption of a sequential process of distance class rejection and acceptation.

* Submitted for publication.

In the description of voluntary movement one often distinguishes decision or initiation from movement execution; this has found its formalization in various theories of open-loop and closed-loop motor control, (see e.g., Stelmach (1976), for a number of reviews) and of intermittency assumptions (Craik, 1947; Vince, 1948) in the execution of other than extremely fast movements. For instance, in the case of a choice reaction task, reaction time RT prior to movement onset may vary with the number of alternative stimuli or responses and with stimulus-response compatibility (Smith, 1968); similarly, in tracking a continuously varying input signal, subjects often tend to adjust their control device with discrete steps in time (Navas & Stark, 1968).

One explanation for the latter phenomenon might be that quantization of the discrepancy between stimulus and response is involved; however, discrete adjustments in tracking are also found when errors prior to readjustment are well above known psychophysical thresholds. Still, it might be that subjects adopt internal criteria of error tolerance; for instance, Mesarovic *et al.* (1973) have proposed a "satisfaction principle" rather than a control-theory biased optimality principle to account for discrete eye movements. Alternatively, one might assume a speed-accuracy trade-off in the evaluation of a particular stimulus under temporal stress, e.g. in terms of some sort of stimulus integration model.

Nevertheless, the most common explanation of the above given distinction is in terms of sampling or intermittency in *time*. Many authors have presented evidence that evaluation of sensory information, decision based upon it, and execution follow each other in a more or less regular fashion, and various theoretical constructs such as single channelness (Welford, 1968) and quantized timing (Michon, 1967) have been proposed as explanatory tools. Thus, the logarithmic relationship known as Fitts's Law (Fitts, 1954):

$$MT = k \ln(2A_0/W) + MT_0$$

between movement time MT, distance to target centre A_0 and target width W has been explained by Crossmann & Goodeve (1963) as the consequence of an intermittent correction process in which corrective steps of constant duration are made. During each step, the last part pA_n of a larger distance A_n is consistently selected. One may note that, although this model has usually been viewed as a control-theoretic model because of the involvement of feedback during movement, it might equally be viewed as serial scanning, self-terminating, information processing model in line with Fitts's (1954) original explana-

tion. As such, a satisfaction principle "within target range" is employed, rather than some classical control-theoretic point optimum criterium. See also Welford (1968).

Wickens (1968), using a force controller in a step tracking task on a visual display, indeed found periodicities in the order of 10 Hz in the velocity profiles of his subjects, which in some cases did show the stepwise decreasing shape to be expected from this serial scanning model. A problem with such models is, however, that the delays in processing visual information may be too large to fall within one elementary corrective step, implying some form of predictive testing of hypotheses "within target range": for instance, Keele (1968) reports typical processing times for visual information of 200 - 300 msec, whereas the corrective steps in Wickens's case lasted only 100 msec.

Crossman & Goodeve (1963) suggested an other explanation: "This is not as startling a conclusion as it may seem, since all previous RT studies supplied an *external command* to carry out the response, whereas in this case the response is to a change in error signal due to the subject's own motion and hence arriving at a known time. It may be well be ignorance of precise timing that produces long RT." (p. 10).

In some cases, one may wonder whether intermittency should be viewed as reflecting *distinct* or "different" rather than *discrete* but similar motor acts: for instance, Langolf *et al.* (1976) recently presented evidence from a microscopic peg transfer task in which subjects typically undershot the target; in a subsequent homing phase, the peg was inserted into its hole. They argued that this homing phase was under visual control, but from their description of the task it would appear that proprioception was involved also: even if the error at homing onset is perceived visually, the execution of the corrective step might be ballistically controlled, by simply moving the peg until it drops into its hole. From Crossman's (1957) findings, one might expect that the transverse error in the initial distance-covering phase is negligible, so that error correction may indeed be confined to moving into a known direction until peg hole attainment.

Mulder *et al.* (1976) have suggested that continuous versus discrete information processing in motor skills is still an open question, since methodological artefacts may be involved. One such an artefact might be that subjects are biased to discrete motor acts since target information is usually available prior to movement onset; little research has been reported where target information is suddenly or gradually made available during movement execution. Therefore, a number of experiments are described in the present paper where target information prior to or during movement execution in a simple, unidimensional step tracking task is systematically controlled. See Kerr (1975) for an experiment where a binary direction choice had to be made during movement.

That target information during movement may be beneficial in addition to prior information has been clearly demonstrated by investigators such as Woodworth (1899), Vince (1948) and Keele and Posner (1968). Usually, movements must last at least 200 - 400 msec if "current visual control", as Woodworth put it, is to have any additional effect, but it is not clear which part of the visual trajectory is or may be used for correction. Thus, it might be the case that distance to the target at a certain instant in time is sampled and used to generate a proper adjustment command; alternatively, perceived speed may be integrated over an initial movement interval and used to generate an estimate of target error. Finally, a comparatively continuous "homing" phase may oppose theories of discrete target approach.

Under normal circumstances, e.g. in the case of movements at common areas with respect to the body, one may expect a proper "map" between sensory information and motor output control. In such cases, predictive control compensating perceptual and central processing delays might be expected to be possible. However, when this coordination is disrupted, as in the case of displacing lenses (Kornheiser, 1976) or movements at uncommon places with respect to the body (Hay *et al.*, 1976), this might be difficult. Still, if the distortion is of a comparatively simple type, e.g. involving only a shift or scaling factor between perceptual input and motor output, predictive control might be reacquired in an adaptive learning process. In a number of experiments described below, however, such adaptation is necessary during each individual trial.

One cause for intermittency might be of a sensory nature: it is well known that the eye has a rather low bandwidth, and that perception is strongly reduced during saccadic eye movements. It might, therefore, be possible that higher processing rates, perhaps even of a virtually continuous nature might be found for faster sensory modalities; audition would seem to be a good candidate for this purpose.

Apparatus and Task

Subjects sat at a table and rested their right elbow on a cushion. They held an IR power-LED in their right hand and oriented it toward a SELSPOT camera (Lindholm, 1973; Woltring, 1974) which observed the LED's direction. The 2-dimensional coordinates of the LED's image were AD-converted in the SELSPOT central processing unit at 9-10 bits accuracy per axis and transmitted by wire to a PDP-11/45 computer in an adjacent room (RT-11 operating system, floating point software) under direct memory access via a DR11-B interface. Feedback was derived from the horizontal camera output only and presented to the subject in auditory or visual form. Auditory feedback was pitch-controlled, and consisted of square-waves with an intensity well over auditory

thresholds; pitch was proportional to image distance between the subject's hand and some reference point. Square waves were chosen in order to avoid problems with subauditory l.f. signals which were found to occur for sinusoidal or triangular signals when subjects approached the reference position. Visual feedback was presented as a light-spot moving over the screen of a VT-01 display operated in non-storage mode. Here, the reference point was indicated by a stationary light-spot on the screen.

Subjects had to move the light source in their hand in the coronal plane normal to the camera axis, by rotating about their elbow in such a way as to attain the reference signal. If overshoot occurred, pitch would increase again, and the subject's light spot would "bounce" at the reference lightspot. In all cases, a linear relationship was maintained during trials between horizontal camera output and feedback; however, the gain factor $G = d(\text{feedback})/d(\text{distance covered})$ could be varied or kept constant between trials in an experimental session, depending on the experimental requirements. The movement plane was at 1.70 m from the front nodal point of the camera, which was equipped with a 5 cm focal length lens, allowing a width of field of 85 cm.

Image distortion due to distance variations between LED and image plane or to lens errors (Woltring, 1975, 1976) were not corrected, since the errors of the human operator within the sagittal plane appeared to be much larger. Similarly, the non-linear relation between horizontal displacement sensing and angular lower arm rotation was neglected, subjects being informed that horizontal displacement only was detected. For some subjects, the camera was slightly tilted about the optical axis as to balance abducent and adducent arm movement range with respect tot the point of symmetry at the optical axis.

In the case of visual feedback, the reference light-spot on the display was at the left edge (as seen from the subject) if he had his hand to the right of the optical axis at trial onset, and at the right edge in the reverse case. In this fashion, S-R compatibility was improved. The display was to the left of the subject (see Fig. 1), at approximately 1.20 m from his vertical axis of symmetry. The subject was left free to orient himself at will, and room illumination was kept low as to prevent disturbing reflections at the display screen.

Fig. 1. Experimental set-up

Each trial was preceded by an auditory warning signal followed by a randomly varying waiting interval ranging between 1 5 and 2 5 sec (uniform distribution) Subsequently, feedback was turned on and the subject could start his movement (he was instructed not to move before feedback was turned on) The LED's position was then sampled for 1 55 sec, or 500 SELSPOT observations, and the warning signal for the following trial occurred 0 5 sec later Depending upon the experimental treatment, feedback was turned off either at specific movement phases or after the 1 55 sec interval The visual reference point remained visible during the full 1 55 sec interval in the case of visual feedback Each experimental session consisted of 10-25 series at 12-20 trials each The pause between series was usually 20 sec, but could be modified at the subject's wish One session usually lasted 20-45 min

Starting position for each trial was usually the end position of the previous trial, the subject being instructed to move away from the optical axis if he had stopped too closely to it during the previous trial Random distances to be covered were generated, and subjects had to move to the reference position in a fast or slow fashion as instructed The distance distribution was discrete and uniform, with mean value DAV, range DRANGE, and number of equidistant values NRANGE $\geqslant 5$ (subjects reported afterwards to have no idea about this number of distance alternatives)

Depending upon the experimental treatment, initial feedback (pitch, lightspot distance) could vary or remain constant between trials during a session (*a priori* target information or INITV, no *a priori* information or INITF), in addition, feedback could be presented over fixed or varying distances between trials within a series (single and mixed feedback, respectively) Thus, in the case INITF, target information became available during movement only, with the gain factor G varying inversely with the randomly generated distance D to be covered, whereas G was constant in the case INITV, allowing the subject to generate a complete motor program since distance to be covered was known *a priori* in terms of pitch or light-spot positions Initial pitch could vary between 800 and 2000 Hz, and initial distance between the display light-spots ranged between 10 and 18 cm

In all cases, a randomized partial reinforcement schedule of 30% was employed, during reinforcing trials, feedback remained on during the complete trial In addition, each series was preceded by two similar warming-up trials, the first series of a session containing only complete feedback trials By adjusting any initial step until reference-point attainment, subjects could learn the relation between movement and feedback

Data Processing

After each trial (except for the two initial warming-up trials per series), the raw position observations and various trial parameters were stored on a cartridge disk, for processing after session termination All not reinforced trials were processed through Fast Fourier Transform (FFT) methods (Oppenheim & Schafer, 1975, Rabiner & Gold, 1975), with low-pass filtering and differentiation occurring in the frequency domain In some cases, the calculated accelerations were averaged per required distance class and per direction, with the initial acceleration maximum as the common reference point in time, since this quantity turned out to be quite uncorrelated with most experimental treatments within a session Furthermore, mean values, standard deviations and cross-correlations for various quantities were evaluated

Since floating-point hardware was not available, signal processing was conducted on integer data, with scaling measures to avoid overflow and underflow (a procedure technically known as *block floating point*, Oppenheim & Schafer, 1975). In this manner, two trials could be simultaneously analysed, one in the "real" data array, and one in the "imaginary" array (rounding errors were negligibly worse than for floating-point FFT, and time savings were considerable processing 2 trials in integer mode with an ASSEMBLER language FFT-routine required 2 sec, excluding curve displaying, while use of a FORTRAN-IV floating-point mode FFT-routine required 13 sec for the same two trials, for floating point hardware, processing time is expected to fall below 0 75 sec). By sinusoidal windowing applied to beginning and end of the raw observations, boundary effects in the FFT-procedure were avoided Data were low-pass filtered prior to differentiation, the frequency response of the filter having a sinusoidal shape (half period) with cut-on at 2 Hz, and cut-off at 20 Hz - for larger frequencies, acceleration became too noisy, although velocities remained quite acceptable. The regression data to be discussed below were rather insensitive to small changes in these frequencies, in the order of 0 03 changes in correlation coefficients for 2 Hz changes, and $df > 70$

After some pilot studies, start and end points of a "ballistic" step were defined as follows the time-instant T_3 of maximal velocity (zero acceleration) is sought, and the time instants T_2 of maximal acceleration prior to T_3 and T_4 of maximal deceleration after T_3 are established Movement start instant T_1 is defined as the last zero crossing of velocity or acceleration, whichever is the later one, prior to T_2, and movement step end is defined to occur at the first zero crossing (T_5) of velocity or acceleration after T_4, whichever is earlier, see Fig 2. In this manner, manual tremor effects prior to movement initiation were minimized, and clear additional steps after the initial step movement were removed from analysis. It would appear, though, that more complicated step identification algorithms are necessary The distance covered D' is defined as the filtered distance difference between the time instants T_5 and T_1, and this quantity is related to the required distance D through linear regression analysis.

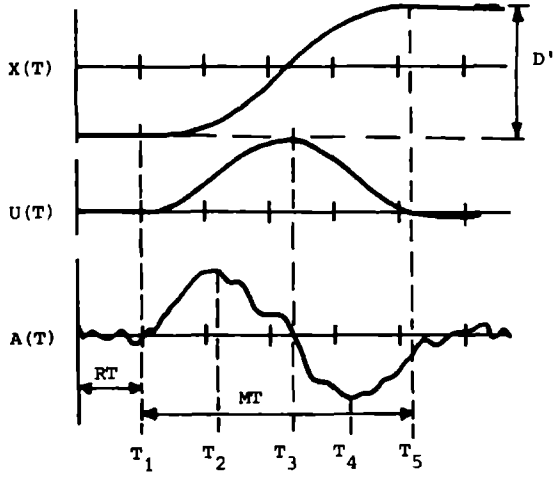


Fig. 2. Definition of time-instants during a step movement and of the distance D' covered (x(t) displacement, v(t) velocity, a(t) acceleration. Horizontal scale. 200 msec/division, vertical scale arbitrary).

In those cases where tracking performance was limited, there was a clear tendency to move to the mean value DAV of the generated distance distribution during a session. This mean position is of little interest; the most important variable is the *slope* of the regression line which indicates to what extent target information is employed, as illustrated in Fig. 3. In the various figures below, these slopes and their standard deviations as estimated are plotted as a function of the feedback interval FB from movement onset, at which feedback disappears.

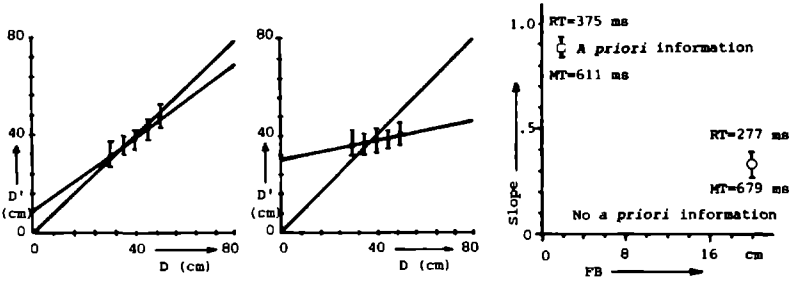


Fig. 3. Ideal and true regression lines, means and standard deviations in distance reproduction for various values of the required distance D as generated at random (left: *a priori* target information (INITV), middle: no *a priori* target information (INITF)). Right: estimated slopes and standard deviations plotted as a function of the feedback interval FB, together with mean reaction and movement times.

Data were analysed in terms of linear regression models, assuming normality of the error distribution in the dependent variables. Based on pilot experiments in which scatter diagrams of the actual distances covered (D') were plotted against the required distance D, this assumption has appeared reasonable. Similarly, the homogeneity and linearity assumptions have been tested by visual inspection only, by viewing regression data such as shown in Fig. 3; an exception is discussed for Experiment II below. Responses deviating by more than 25 cm or 75% from the required distance (100% for subject M2 because of his more extensive training) were discarded.

Most conclusions in the following have been based on differences between regression slopes, rather than on significance of a regression term with respect to some specific and given value (cf. Fig. 3); for this purpose, the Student's test devised by Fisher (1922)[1] has been employed.

Since the degree of freedom is always larger than 70, and usually ranges between 100 and 350, slope estimation may be judged on significance by visual inspection, by assuming normality. For instance, in the case of equal σs and just not overlapping $\pm 1\sigma$ ranges, the z-score amounts to $\sqrt{2}$, and the estimated slopes are significantly different at the 7.9% level; if these ranges are 1σ apart, $z = 3/\sqrt{2}$, and the significance level decreases to 1.3%

Subjects were undergraduate and postgraduate, male and female students and university assistents aged between 22 and 34 years, the students were either paid a nominal remuneration of Hf 3,- per hour, or they fulfilled a course requirement Due to the variability within and between subjects, and their low motivation to participate for more than a few hours, a more extensive series of experiments was run on only one subject (M2). Considering the high variability between subjects, no averaging of data has been conducted over them, rather, the results below highlight various individual differences.

Some of these were quite manifest one subject (M3) was much faster and more consistent in abducent movements which he attributed to his backhand in table-tennis, and an other subject (F3) was blind in the right visual (ipsilateral) field if target information became available to the right of the current fixation point, tracking performance was greatly reduced. Most subjects showed individual movement times for the instruction "move fast and yet accurately", despite oral instructions from the experimenter to increase or decrease speed slightly All but one subject were right-handed, however, the left-handed subject had been trained to use his right hand in writing, and left hand performance appeared inferior to right hand performance during a few pilot sessions All subjects reported normal hearing, with the exception of subject F3, all had normal or corrected vision.

Experiment I

Six subjects (3 males, M1-M3, 3 females, F1-F3) were given pretraining ranging between 50 and 100 trials prior to proper experimental sessions. Sessions of the type INITV, where target distance information is available prior to

[1] For two linear regression models $y = a_i + b_i x + e, i = 1,2, e = N(0,\sigma^2)$, and least squares estimates $\hat{b}_i$ for the slopes b_i, the difference $\hat{b}_1 - \hat{b}_2$ is normally distributed, with mean $\hat{b}_1 - \hat{b}_2$, and variance

$$\sigma_d^2 = \sigma^2 [S_{x1}^{-1} + S_{x2}^{-1}]$$

$$S_{xi} = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

n_i = number of observations

$\bar{x}_i$ = mean value of independent variables $\{ x_{ij} \}$

The difference between the estimated slopes may be compared with a specific test value $b'_1 - b'_2$ by the following student's test

$$t = [(\hat{b}_1 - \hat{b}_2) - (b'_1 - b'_2)] / [s_d^2 (S_{x1}^{-1} + S_{x2}^{-1})]^{1/2}$$

$$s_d^2 = [(n_1 - 2) s_1^2 + (n_2 - 2) s_2^2] / df$$

$$d_f = n_1 + n_2 - 4$$

s_i^2 = residual error variances

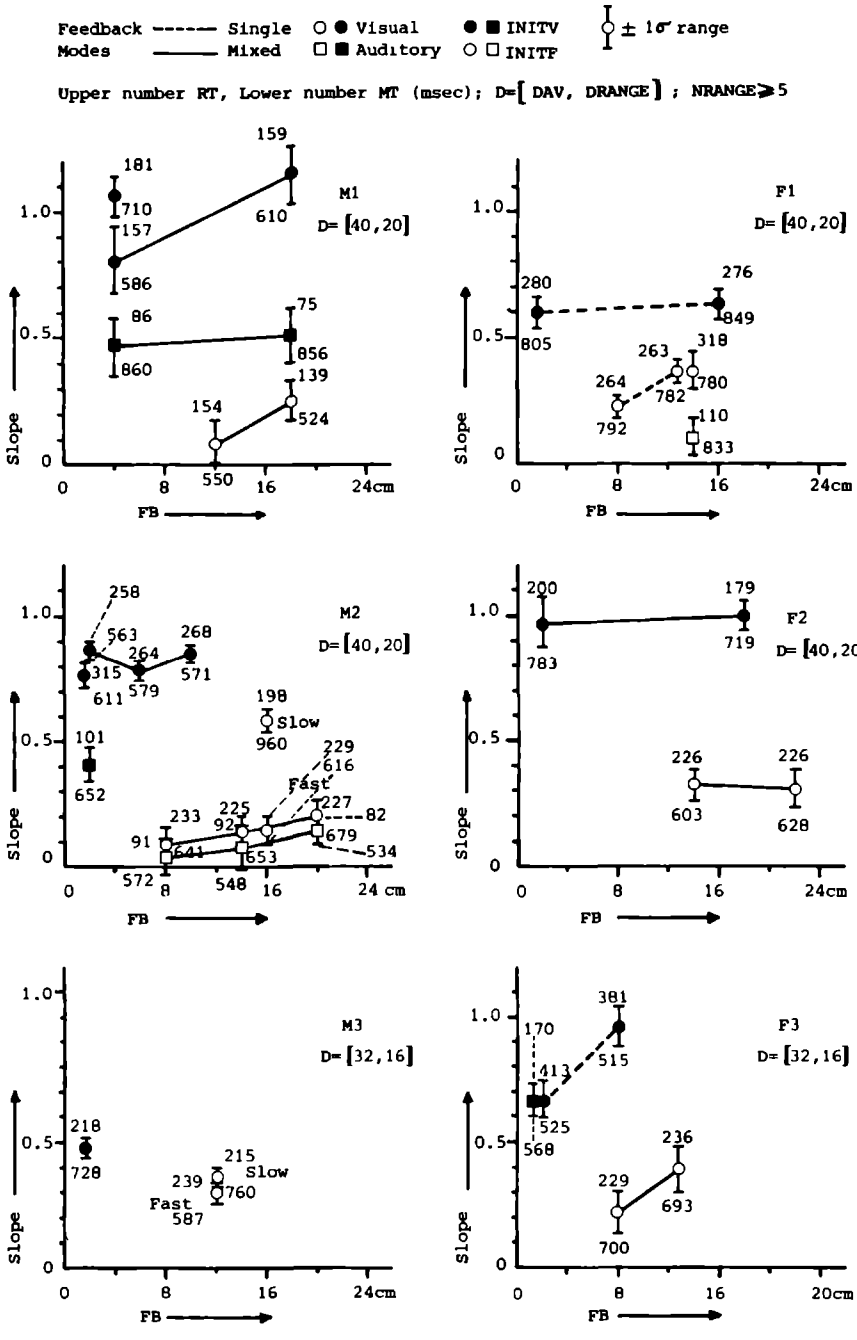


Fig. 4. Regression slopes, RT and MT for the six subjects in experiment I.

movement onset, preceded those of the type INITF, where target distance must be derived from feedback during movement In pilot studies it was found that performance on auditory feedback was inferior to performance on visual feedback. Therefore, sessions with visual feedback preceded those with auditory feedback Care was taken to balance experimental treatments in order to minimize influence of learning and fatigue, either by repeating sessions in reverse order and averaging corresponding sessions, or, in the case of limited subject availability, by choosing treatment ordering in such a way as to render learning counteractive to expected differential effects between treatments

The slope of the various regression curves, together with their standard deviations, mean reaction times RT and mean movement times MT (upper and lower numbers, respectively, in msec) have been plotted in Fig. 4, as a function of the feedback distance from movement onset FB, for the six subjects individually

Auditory versus visual feedback

As apparent from these data, auditory feedback for both prior and current target distance information is inferior to visual feedback, for 3 of the 4 subjects tested on this issue, despite the shorter RT and longer MT found in

Table I. Difference between auditory and visual feedback

Subject	FB cm	SLOPE				Student t p(%)	REMARKS
		Auditory		Visual			
		INITV	INITF	INITV	INITF		
F1	14		0.10		0.37	< 10	
F3	2	0.67		0.67		–	
M1	2	0.47		0.81		< 5	mixed feedback
	18	0.51		1.15		< ½	
M2	2	0.41		0.77		< ½	
	10 14 18		0.09		0.15	< 25	mixed feedback, averaged

cases for auditory feedback. One subject (M1) overcompensated in the case of prior and current visual information.

Some statistical tests are presented in Table I. For subject M2, the difference between auditory and visual feedback (mixed feedback over 8,14 and 20 cm) was insignificant in the no prior information case, and his regression slopes are quite small. For subject F3 only, auditory and visual prior information were equally effective. Possibly, her right field-of-view blindness has entailed extensive practice in relying on auditory cues, a level of practice which could not be attained for the other three subjects during experiments.

Subjects generally reported that auditory feedback was rather strenuous as compared to visual feedback, one subject even giving up after some pre-training. For this reason, it was decided to rely further on visual feedback only.

Prior target distance information (INITV)

Considering the effect of current feedback during the initial movement interval *in addition* to prior distance information, mixed results have been found. Thus, for a single feedback-distance within sessions, subject F1 does not improve, whereas subject F3 does substantially ($p<5\%$), despite her shorter MT. In the mixed feedback-case, subjects M2 and F2 do not significantly improve with increasing FB, whereas subject M1 does ($p<5\%$). Apparently, subjects F3 and M1 were quite able to correct an initial ballistic move-

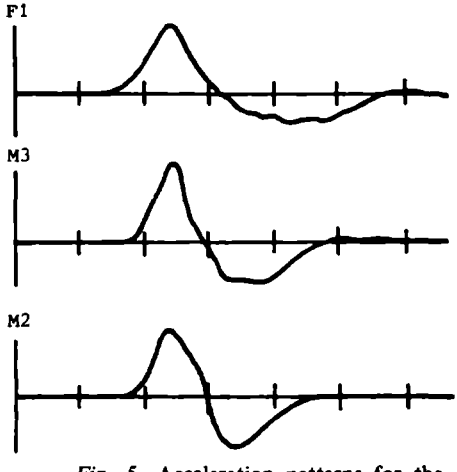


Fig. 5. Acceleration patterns for the subjects F1, M3 and M2 (*a priori* information case, visual feedback, FB = 1-2 cm; data averaged for one distance class, with the acceleration maximum as common reference point (N = 10-15)).

ment by evaluating current feedback during initial movement execution, rather than through terminal error assessment at the end of the initial step D'. All subjects, however, were capable of learning the visual step tracking task with prior information, exhibiting regression slopes > 0.4, and their data suggest individual upper performance limits for those experiments where no prior information is available (INITF, to be discussed below).

The control strategy of the step appeared to vary between subjects as is apparent from various correlation coefficients with the distance D to be covered, for the *a priori* information case under visual feedback. Thus, subject M1 clearly modulated his peak deceleration value (ρ = -31%) and delay (ρ = 39%), for fairly constant initial acceleration values between distance classes, whereas subject M2 modulated both the acceleration amplitude (ρ = 39%) and deceleration amplitude (ρ = -33%) for fairly constant timing between these control pulses. The other subjects did not exhibit sufficiently high correlation coefficients to decide upon any of such or other strategies.

The relative durations and shapes of the acceleration and deceleration impulses varied considerably between subjects; for instance, subject F1 and M3 showed relatively extended, low deceleration phases and rather strong acceleration phases, whereas subject M2 exhibited rather symmetrical start and stop impulses: see Fig. 5.

No prior target distance information (INITF)

It was found for all subjects that performance in terms of regression slope values was greatly reduced if no prior target distance information was made available, as apparent from Fig. 4. In order to establish to what extent temporal limitations were involved, two subjects were instructed to move fast or slowly, in separate experimental sessions: see Table II. Only in the case of subject M2, movement time had a significant effect on distance reproduction; however, this subject showed a much larger increase in mean movement time.

Table II. Influence of fast/slow instructions on distance reproduction and mean movement time, for visual feedback without prior distance information

Subject	FB cm	Fast		Slow		Slope equality test
		Slope	MT (msec)	Slope	MT (msec)	
M2	16	0.15	616	0.58	960	$p < ½\%$
M3	12	0.30	587	0.36	760	$p < 25\%$

In a similar fashion as in the *a priori* information case (INITV), different effects were found between subjects for mixed feedback treatments. Thus, slight improvement with increasing FB, i.e., positive slope *rate*, was found for subjects M1, M2 and F3, but not for subject F2. This effect was investigated more extensively with subjects F1 and M2, as shown in Fig. 6[2]. For subject

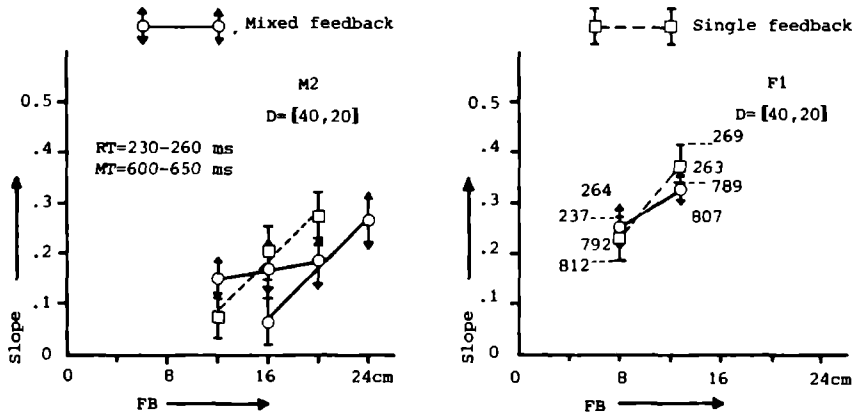


Fig. 6. Influence of feedback range and mixed versus single feedback in visual tracking without *a priori* target distance information.

M2, who was treated most extensive, it was found that slope rate was significantly positive for single feedback sessions over the feedback interval FB = 12-16-20 cm, whereas it was not for mixed feedback over the same FB-values. However, for the larger values FB = 16-20-24 cm, mixed feedback did entail a significantly positive slope rate. Similarly, subject F1 showed slope rates which were significantly positive, both for mixed and single feedback (FB = 8-14 cm), but which did not differ significantly from each other, although slope rate was slightly larger in the single feedback case; see Table III.

[2] Here, more than 2 slope values are to be tested on significance of their differences. Rather than using the test of the previous footnote, one might also test on significance of the slope rate with respect to zero, by fitting a regression line to the previously estimated slope values as dependent, and with the FB-values as independent values. Since the degree of freedom per slope estimate is large (df $>$ 90), each of these may be viewed as a normally distributed, stochastic variable with known standard deviation equal to the estimated value. By means of ordinary least squares analysis, the slope rate then follows as a weighted least squares estimate with known standard deviation, to be tested with respect to zero by evaluating the z-score. Alternatively, the difference between slope rates may be tested on significance in a similar fashion as before.

Table III Influence of feedback range and mixed versus single feedback in visual step tracking without *a priori* target distance information

Subject	FB-interval cm	Feedback mode	Slope rate cm^{-1}	Significance tests	
				zero?	equal?
M2	12-16-20	Single	2.52×10^{-2}	$p < ½\%$	$p < 5\%$
		Mixed	4.35×10^{-3}	–	$p < 5\%$
	16-20-24	Single	$2\,55 \times 10^{-2}$	$p < ½\%$	
F1	8-13	Single	3.04×10^{-2}	$p < 1\%$	$p < 25\%$
		Mixed	$1\,60 \times 10^{-2}$	$p = 10\%$	

The correlation coefficients of peak acceleration and velocity (values and instants) with the required distance D were insignificantly different from zero, indicating that subjects moved to some average distance value, usually close to the mean distance DAV, only movement time was significantly correlated ($\rho_{crit} = 25\%$) Clearly, an initial step was corrected from approximately the maximal velocity instant T_3

Acceleration profiles for individual trials often showed an additional step or periodicities in the order of 10 Hz during the deceleration phase, usually

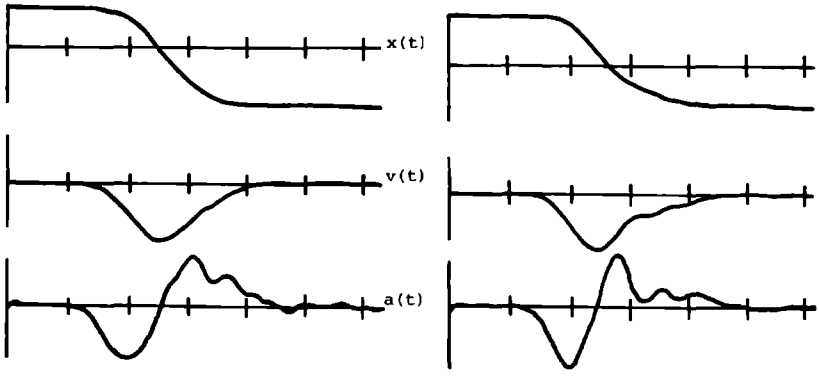


Fig. 7 Individual movement profiles for short and long distances within a session, for visual feedback without *a priori* distance information x(t) displacement, v(t) velocity, a(t) acceleration. Horizontal scale 200 msec/division

after the deceleration maximum T_4, suggesting intermittent control of the final adjustment phase see Fig 7 as an example Such periodicities will tend to cancel in averaging procedures (cf Fig 5), since there is no reason to assume that peak acceleration or other reference points during the initial ballistic and intermittency in homing onto the target are correlated. Perhaps, averaging the amplitude spectrum of an ensemble might reveal typical intermittency components Only for the largest distance classes, evidence of additional correction or intermittency became sometimes apparent from averaged acceleration data see Fig. 8.

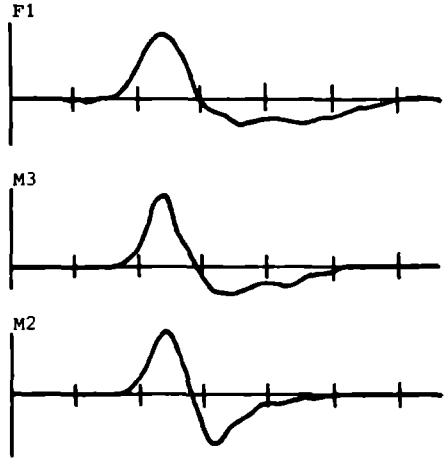


Fig. 8. Step adjustment and intermittency in averaged acceleration patterns, for subjects F1, M3 and M2 (no *a priori* information, visual feedback, data averaged with the acceleration maximum as common reference point, N = 10-15).

Introspectively, the last part of the feedback interval and the position on the display where the feedback signal disappeared seemed an important cue in the generation of a corrective command In order to verify this impression, the following, modified experiment was designed. Feedback was presented until movement onset, as defined by a distance change of 1 cm, and interrupted until ½FB distance had been covered Subsequently, feedback was returned until FB distance had been covered, and the regression slope for this experiment was compared with the slope obtained for feedback since movement onset The data in Figure 9 and in Table IV clearly show that also feedback

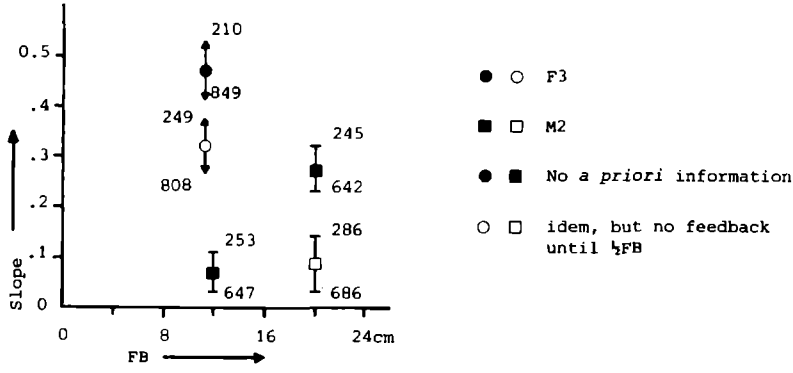


Fig. 9. Integration of current feedback: regression slope values for full feedback between 0 and FB cm from movement onset, and for feedback during the last half of the interval FB.

during the first half of the movement feedback interval FB has a significant influence on distance reproduction. Thus, even if the position at which feedback disappears on the display is of importance, its influence is limited as regards the distance D' defined between the time instants T_1 and T_5: this is to be expected since it occurs at the shortest time interval prior to T_5, and may well fall within a visual reaction time.

Table IV. Integration of current feedback: regression slope values for full feedback between 0 and FB cm from movement onset, and for feedback during the last half of the interval FB

Subject	FB-interval	Slope	Significance tests	
			Zero?	Equal?
M2	0 – 12 cm	0.072	p = 5%	p < ½%
	0 – 20 cm	0.272	p < ½%	p < ½%
	10 – 20 cm	0.071	p = 5%	
MG	0 – 11½ cm	0.472	p < 5%	p < 5%
	5¾ – 11½ cm	0.319	p < ½%	

In the no prior information case, distance information became available only gradually. A different situation arises if target information is made fully available once a certain distance has been covered, and some experiments have been done after modifying the experimental paradigm of the previous section as follows.

The light spot corresponding to the subject's right hand always started at the edge of the display, and the reference point at trial onset was given at 7 cm from the other edge. Feedback remained available over a distance FB (for non-reinforcing trials), after which it was removed. At the same time, the reference point jumped to a new and final position to the left or right of the initial one which corresponded to the average distance DAV, and subjects had to attain the new reference position as fast and accurately as possible. In all other aspects, task and configuration remained the same as previously, the gain factor G being constant between trials within a session.

The data in Fig. 10 show regression slopes of distance reproduction as before, for 3 subjects, together with comparative data for *a priori* information from Experiment I. Clearly, adjustment deteriorates with increasing distance FB at which the target assumes its new position; see also Table V.

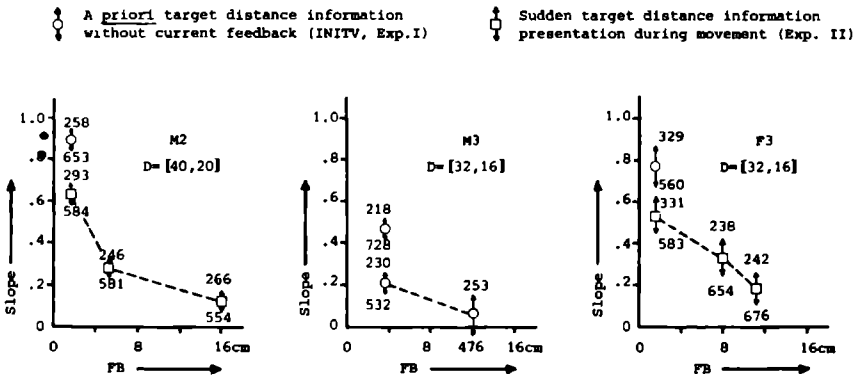


Fig. 10. Distance reproduction for prior and current, sudden distance information presentation.

Table V. Distance reproduction for prior and current, sudden distance information presentation

Subject	Experiment	FB cm	Linearity test	Slope		
				Value	Significance test	
					Zero?	Equal?
M2	I - *a priori* inform.	1 6	–	0 893	p < ½%	p < ½%
	II	1.6	p < ½%	0.622	p < ½%	p < ½%
		5 2	p < 2½%	0.277	p < ½%	p < 1 %
		16	p < 10 %	0.199	p < ½%	
F3	I - *a priori* inform.	1.6	–	0 769	p < ½%	p < 5 %
	II	1.6	p < 25 %	0.542	p < ½%	p < 5 %
		8	p < 2½%	0.331	p < ½%	p < 15 %
		11	–	0.180	p < 2½%	
M3	I - *a priori* inform.	1 6	p < 25 %	0.481	p < ½%	p < ½%
	II	1.6	p < 25 %	0.211	p < ½%	p < 12 %
		12	–	0	–	

However, the confidence levels for the linearity hypothesis show that the validity of the present regression model deteriorates for this second experiment, with 2 of the 3 subjects Typically, distinct, linear regression models could be fitted for distances D smaller and larger than average (D<DAV and D>DAV, respectively) In a few cases, such effects have also been observed for experiment I, but not as clear as in the present experiment Therefore, the data for subject M2 were further analysed, since he showed most consistent behaviour and the shortest movement times, having had most experience of the three subjects

The regression data in Fig 11 show that subject M2 clearly differentiated distances shorter and longer than the average (and initially presented) distance

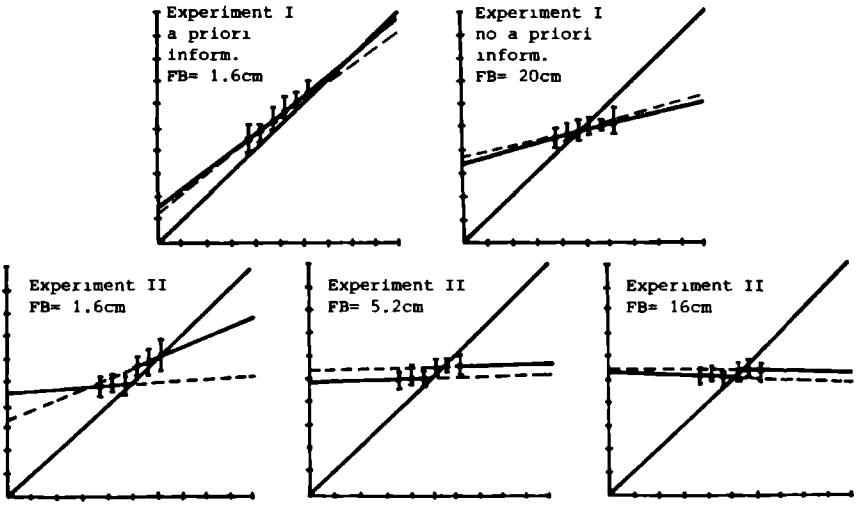


Fig. 11. Regression lines for D < DAV and for D > DAV; subject M2, both directions combined. Horizontal and vertical scales 8 cm/division.

DAV. Only in the case of the shortest FB-value (1.6 cm), the regression slope for D > DAV differed significantly from zero (0.678, $p < 1\%$) and from the slope for D < DAV ($p < 1\%$). After classing these data into adducent and abducent movements, it appeared that the above effect was wholly due to adducent movements, since the two regression lines for abducent movements were both virtually horizontal and differed insignificantly in slope: see Fig. 12. For other movements, no significant differences between abducent and adducent movements were found.

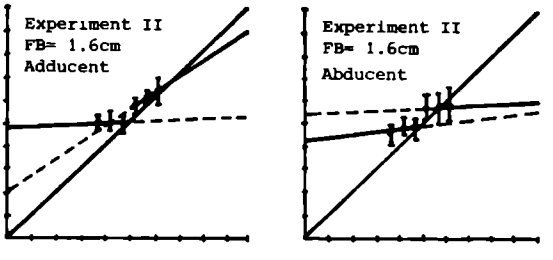


Fig. 12. Regression lines for D < DAV and for D > DAV; subject M2, for adducent and abducent movements; FB = 1.6 cm.

Table VI. Differences between regression lines for D < DAV and D > DAV; subject M2 (MT = 550-560 msec)

Type experiment	FB cm	Movement direction	Regression slope					Estimate at DAV		
			D < DAV		D > DAV			value (cm)		
			value	zero?	value	zero?	equal?	D < DAV	D > DAV	equal?
I - *A priori* inf.	1.6	both	0.766	$p < ½\%$	0.834	$p < ½\%$	–	43.0	42.7	–
I - no *a priori* inf.	20	both	0.305	$p < 2½\%$	0.237	$p < 5\%$	–	40.3	40.3	–
II	1.6	adducent	0.678	$p < ½\%$	0	–	$p < 1\%$	40.5	44.5	$p < 1\%$
		abducent	0	–	0	–	–	39.2	46.2	$p < ½\%$
	5.1	both	0	–	0	–	–	40.6	44.6	$p < ½\%$
	16	both	0	–	0	–	–	40.7	41.9	$p < 25\%$

Significance tests on these effects are presented in Table VI; unlike the case in Experiment I, both slopes and regression intercepts have to be compared[3]. Since DAV appeared to be the initial and symmetrical distance value being aimed at, a proper procedure seemed to be to test on the difference between extrapolated regression estimates for DAV (which was never presented as final distance value, by choosing an even number of distance classes NRANGE), for the two regression models obtained for smaller and larger D than average.

Thus, in the present experiment, differentiation between distance alternatives was largely of a binary nature; the regression lines for the FB = 1.6 cm adducent movements intersected close to the mean of the D < DAV distance range, suggesting that shorter distances were not differentiated from this point.

Discussion

Auditory versus visual feedback

The hypothesis that a faster sensory channel such as audition would be superior to vision in fast distance differentiation was not supported by the data, except for subject F3 who was blind in the right visual field. The shorter RT to auditory feedback onset should perhaps be explained as an arousal phenomenon (Sanders, 1975) since the signals were rather loud and presented at medium time uncertainty (see also Sanders, 1976). It is tempting to explain this inferiority to visual feedback in terms of stimulus-response compatibility, but this notion does not seem to convey much more than a name for the very effect being observed. Yet, auditory *location* cues such as might be presented through stereophonic stimulation would seem to be a more "compatible" feedback source. If compatibility simply denotes extensive extra-laboratory experience, then pitch clearly is a most incompatible source of information, perhaps excluding professional musicians. None of the subjects belongs to this class.

[3] The variance of any regression estimate $y_i^*(x) = b_i x + a_i$, using the notation of footnote [1], is equal to:

$$s^2_{y_i^*} = \sigma^2\left[\frac{(x - \bar{x}_i)^2}{S_{x_i}} + \frac{1}{n_i}\right]$$

The difference $y_1^* - y_2^*$ may be compared with a specific test value $y_1' - y_2'$ by the student's stochastic:

$$t = [(y_1^* - y_2^*) - (y_1' - y_2')]/\left[s_d^2\left\{\frac{(x - \bar{x}_1)^2}{S_{x1}} + \frac{(x - \bar{x}_2)^2}{S_{x2}} + \frac{1}{n_1} + \frac{1}{n_2}\right\}\right]^{1/2}$$

A priori information case

The high slope values obtained in the visual feedback mode without current feedback (FB = 1-2 cm) indicate that distance reproduction could be rather well learnt by most subjects, although they always underestimated the distance range DRANGE. Only subject M3 exhibited a regression slope < 0.5. Since the *a priori* case was devised mainly to assess upper performance limits, in order to have a standard for the two no *a priori* information cases (Exp. I and II), no catch trials were included to verify whether lower slopes indicate a tendency to respond too fast. After all, in the other cases fast information processing is necessary.

Isserlin (1914) in his investigations of finger movements made a distinction between "motor" and "sensory" movements (cf. Lange, 1888). The former correspond to Donders's (1868) *a-reaction* or simple RT, the latter to Donders's *c-reaction* (no reaction allowed for certain stimuli). He found clear individual differences between his subjects, some revealing a typical "motor" attitude, and some being biased to "sensory" movements. The former tend to rely more on movement preparation than on stimulus differentiation, as apparent from their large number of false movements in the c-reaction task. If these movement classes are generalized to encompass Donders's *b-reaction* (choice reaction task), one would expect that a failure to distinguish movement/no movement stimuli would transfer to failure in distinguishing alternative distance stimuli. This does not necessarily imply that information processing *during* movement would be equally inferior; rather, movement preparation and perceptual differentiation cannot occur simultaneously. In this fashion, the high increase in distance reproduction when current feedback during the initial movement phase is added (subjects M1 and F3) may be explained.

No a priori information case

In a similar vein, high discrepancies between the two cases of *a priori* and no *a priori* information in Experiment I as shown by the subjects M1, M2 and F3 might suggest a bias to "sensory" movements, with much emphasis on stimulus differentiation prior to movement onset, or *response programming* (Schmidt, 1975).

However, low distance differentiation in the no *a priori* information case may also be explained from psychophysical data on acceleration perception (Schmerler, 1976). It is known that visual perception of acceleration in man is rather limited, and this might be an important cue in evaluating initial, current feedback.

The results in Fig. 9 and Table IV suggest that feedback information is integrated over the whole initial trajectory if made available in gradual form

under single-feedback sessions. Interestingly, for the shortest movements such information is terminated rather soon before the initial ballistic passes through the required distance (typically, in the order of 100-150 msec). Since the subject is not capable of preventing overshoot, Crossman and Goodeve's (1963) contention that visual feedback from one's own movements is processed faster due to its arriving at a known instant, is not substantiated. Even more interestingly, the regression lines are essentially straight, with fairly homogeneous error distributions across distance classes in most cases, suggesting that distance reproduction for $D > DAV$ is not better than for $D < DAV$, despite the larger amount of time available. It seems that the present paradigm of gradual distance information accumulation entails a limited amount of continuous goal alignment by the subject, rather than discrete choices between distance alternatives as proposed by Crossman & Goodeve (1963).

One might expect that residual variance or "variable error" would increase with distance without feedback, as reported by Howarth *et al.* (1971) for feedback termination during the deceleration phase. Although residual variance usually increased with decreasing target information (Experiment I: decreasing FB), the error distribution per FB-value is usually constant within sessions. It is possible that a counterbalancing effect is caused by time limitations for shorter movements, inducing subjects to correct earlier for short distances. Alternatively, it might be that subjects assume a more conservative strategy in the case of reduced feedback, compatible with some self-imposed error tolerance for these smaller distances. Whichever the case may be, increased mean movement time may lead to improved distance reproduction as shown for subject M2 in Fig. 4 and in Table II.

The observations on subject M2 (Fig. 6) suggest that for shorther feedback intervals in the mixed-feedback case, a strategy was adopted to rely exclusively on feedback common to all trials. It is as if feedback is integrated from movement onset until $FB_{min} = 12$ cm, and that any additional feedback is discarded. Possibly, the difficulties of adequately perceiving initial acceleration were compensated by focusing on visual input, rather than on visual and proprioceptive input. For larger mixed feedback intervals, differentiation then becomes possible, and subjects process both visual and proprioceptive feedback within trials. For single-feedback sessions, full attention may always be given to visual input since the movement feedback interval remains constant. Thus, higher or equal distance regression slopes per FB-value would be expected for single feedback as compared with mixed feedback. This was not the case for single feedback at FB = 12 cm; an explanation is suggested below in terms of the training procedure adopted throughout all experiments.

The reverse explanation could be used to account for the difference in the mixed-feedback cases at FB = 16 cm. Since much larger feedback intervals

could occur, up to approximately maximum velocity, it might be that initial acceleration was insufficiently monitored in the larger, mixed-feedback case with FB = 16-20-24 cm.

The results of Experiment II are quite different; in view of the data in the Figures 11 and 12 it would seem that subject M2 first differentiated between larger and smaller movements than average (reference point jumping away from or toward the feedback point on the display), and that he would further differentiate if time permitted him doing so. The latter could only be found for the smallest FB-value (1.6 cm), and only for adducent movement, Possibly, visuo-motor pretraining for manipulations in common areas with respect to the body is involved (Hay *et al.*, 1976). Since the regression line for $D > DAV$ and adducent movement (Fig. 12) intersects the other regression line at approximately the mean of the $D < DAV$ distribution, it seems that the subject adopted a strategy of bisecting distance classes in a sequential fashion: initially, he moved to the midpoint of the total distance range, and subsequently to the midpoint of the chosen subdivision. If sufficient time was left, further differentiation was realized. Of course, this explanation is closely related to Crossman and Goodeve's (1963) sequential information processing model discussed earlier.

It is tempting to speculate that stimulus differentiation in the *a priori* case (Experiment I) occurs in a similar manner; this would be in conformance with Hick's Law (1952), where RT varies logarithmically with the number of alternatives, and which has been similarly explained as a sequential process of selecting response alternatives.

Thus, rather different distance reproduction effects have been found, depending upon whether distance information is gradually or suddenly presented during movement execution. For *gradual* information it seems that a single motor *program* is continuously adjusted, possibly with intermittent efferent motor *commands* as suggested before in view of the Figures 7 and 8. Introspectively, it was felt that for larger feedback intervals with gradual feedback, tentative steps were sometimes made based on early current feedback, to be followed by additional corrections once further feedback had been processed. Thus, some limited form of sequential distance class rejection might be involved. Such information processing, in line with Crossman and Goodeve's work, seems clearly to occur if distance information is *suddenly* presented at a known distance FB from movement onset.

Training procedure

One possible criticism concerns the training procedure adopted throughout all experiments. Do subjects indeed easily learn to code prior information or initial, current feedback through partial reinforcement by means of con-

tinued feedback until target attainment? It might be that for those experimental treatments where small slopes were found, subjects exclusively relied on continued feedback during reinforcing trials, thus failing to relate *a priori* information or early current feedback to extended feedback. This seems unlikely if the order of experimental treatments is considered, since sessions with larger FB-values commonly preceded those with lower FB-values (Experiment I), thus allowing some "chaining" of feedback intervals across sessions. In the mixed-feedback case, such chaining should occur within sessions. It might be, however, that such hooking-up of feedback intervals failed in some cases, especially with regard to the data in Fig. 6 and in Table III.

Presentation of KR (knowledge of results) as a graphic or numeric error value after movement termination might improve learning. Alternatively, some *adaptive* tracking technique (Kelley, 1969) might be a promising approach. Here, a moving average of errors over trials is used to control the feedback interval, by comparing it with a tolerance value. If larger, FB is increased (Experiment I), if smaller, FB is decreased. By slowly reducing the error tolerance, subjects might gradually decrease their feedback requirements and yet maintain high targeting accuracy. Movement time might be controlled in a similar way, by manipulating trial duration. However, such procedures are likely to entail instabilities or oscillations in FB, if the rate of error tolerance change is not carefully controlled.

Concluding Remarks

Various authors (e.g., Megaw, 1974; Pew, 1974; Schmidt, 1975) have suggested that motor control is a hierarchical process; thus, general monitoring routines are called upon depending on the task to be performed, calling in turn upon more elementary movements which are intertwined in a coordinated manner, reminiscent of the structure of computer routines passing parameters to subroutines which subsequently take over control (Kerr, 1976). Unlike most contemporary digital computers, however, man and animal may process information simultaneously at several levels; these levels may interrupt each other if the need arises to allocate more "processing capacity" to one particular level. On such a model, some old questions such as the storage and novelty problems (Schmidt, 1975) in detailed descriptions of complex motor sequences may be resolved: rather than storing a large list of output "values", and selecting the required one by scanning for the given argument, the output value is algorithmically calculated - for instance, by interpolation from a few reference values.

Mountcastle (1975) recently presented some physiological evidence on "generalized motor programs" which are independent of target direction and movement trajectory. He interpreted his results in terms of the *holistic* or "Gestalt" aspects of the anticipated movement, a notion which corresponds to Bernstein's (1967) *topology* of movement, and to Van Galen's (1974) *focal* information processing. These aspects are dual to the *metric* (Bernstein) or *ambient* (Van Galen) movement aspects which pertain to geometrical aspects such as direction and extent.

It is suggested that gradual distance information presentation (Experiment I, no *a priori* information) corresponds to continuous adjustment of the "input parameters" of a single routine or "motor programme", and that the sudden information presentation case *during* movement (Experiment II) corresponds to sequencing a series of repeated or new subroutine calls; hypothetically, the occurrence of a digital decision "light-spot left or right" during the initial movement phase has called for a routine which cannot but implement discrete choices in later movement phases.

Acknowledgements

The author would like to thank Mr. W. Klink for the drawings in this paper, Mr. R. Gras for photograph processing, and Mrs. J.A. Thomassen-Baker for linguistical suggestions on the final manuscript. Furthermore, he would like to acknowledge some interesting discussions with Dr. J. Adams and Dr. Ch.D. Wickens at the University of Illinois at Urbana-Champaign, during a recent visit to Canada and the United States of America which was made possible in part through a travel grant from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

References

Bernstein, N.: The Coordination and Regulation of Movements (Pergamon Press, Oxford 1967).

Craik, K.J.W.: Theory of the human operator in control systems. I - The operator as an engineering system. Brit. J. Psych. *38*(1947)56-61.

Crossman, E.R.W.F.: The speed and accuracy of simple hand movements. In: E.R.F.W. Crossman and W.D. Symons: Report to MRC and DSIR Joint Committee on Individual Efficiency in Industry (quoted in Welford, 1968).

Crossman, E.R.F.W., and Goodeve, P.J.: Feedback control of hand-movement and Fitts's Law. Commun. Experimental Psychology Society (1963) (quoted in Welford, 1968).

Donders, F C Die Schnelligkeit psychischer Prozesse. Arch. Anat.Phys.(1868) 657-681.

Fisher, R.A.. The goodness of fit regression formulae and the distribution of regression coefficients. J.R Stat. Soc. *85*(1922)597 (also contained in R.A. Fisher. Contributions to Mathematical Statistics, Wiley, New York 1950).

Fitts, P.M. The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psych. *47*(1954)381-391.

Hay, L , Brouchon-Viton, M., and Rabattu, M. L'ajustement visuo-moteur chez l'homme - recherche de certains facteurs critiques du mouvement. Le Travail Humain *39*(1976)1,53-62.

Hick, W.E : On the rate of gain of information. J. Exp. Psychol. *4* (1952)11-26

Howarth, C I , Beggs, W.D.A., and Bowden, J.M . The relationship between speed and accuracy of movement aimed at a target. Acta Psychol. *35*(1971)207-218.

Isserlin, M Ueber den Ablauf einfacher willkurlicher Bewegungen. Psych. Arb. *6*(1914)1-195.

Keele, S.W Movement control in skilled motor performance. Psych. Bull. *70*(1968)6I, 387-403.

Keele, S.W. and Posner, M.I.· Processing visual feedback in rapid movement. J. Exp. Psychol. *77*(1968)155-158.

Kelley, Ch.R. The measurement of tracking proficiency. Human Factors *11* (1969), 43-64.

Kerr, B. Processing demands during movement. J. Mot. Beh. *7*(1975)1,15-27.

Kerr, B Decisions about movement direction and extent. J. Hum. Movem. St. *3*(1976)199-213.

Kornheiser, A.S.. Adaptation to laterally displaced vision: a review. Psych. Bull. *83*(1976)5,783-816.

Lange, L. Neue Experimente uber den Vorgang der einfachen Reaktion auf Sinnesreize. Philos. Stud. *4*(1888)479-510.

Langolf, G.D. Chaffin, D B., and Foulke, J.A.: An investigation of Fitts's Law using a wide range of movement amplitudes. J. Mot. Beh. *8*(1976)2,113-128.

Lindholm, L.-E An optoelectronic instrument for remote on-line movement monitoring. In Nelson & Morehouse (Eds) Biomechanics IV, 510-512 (University Park Press, Baltimore 1974).

Megaw, E.D.. Possible modification to a rapid on-going programmed manual response. Brain Res. *71*(1974)425-442

Mesarovic, M D., *et al.* Satisfaction principle in modelling biological functions. Kybernetes *2*(1973)67-75.

Michon, J.A. Timing in Temporal Tracking (Van Gorcum, Assen, 1967)

Mountcastle, V.B. The view from within - pathways to the study of perception. The Johns Hopkins Med. J. *136*(1975)109-131.

Mulder, G., Michon, J.A., and Moraal, J. Motorische vaardigheden. In. J.A. Michon, E.G.J. Eijkman and L F.W. de Klerk. Handboek der Psychonomie, ch. 17 (Van Loghum Slaterus, Deventer 1976).

Navas, F., and Stark, L.. Sampling or intermittency in hand control system dynamics. Biophys. J. *8*(1968)251-302.

Oppenheim, A V. and Schafer, R.W : Digital Signal Processing (Prentice-Hall, Englewood Cliffs, N.J. 1975).

Pew, R W. Human perceptual-motor performance. In B H. Kantowitz (Ed.)· Human Information Processing - Tutorials in Performance and Cognition, ch. 1 (Lawrence Erlbaum Ass., Hillsdale, N.J. 1974).

Rabiner, L.R. and Gold, B. Theory and Application of Digital Signal Processing (Prentice-Hall, Englewood Cliffs, N J. 1975)

Sanders, A.F The foreperiod effect revisited Quart J. Exp. Psychol *27*(1975)591-598

Sanders, A F Structural and functional aspects of the reaction process In S. Dornic (Ed.)· Attention and Performance VI. Acad. Press, New York (in print)

Schmerler, J. The visual perception of accelerated motion Perception *5* (1976) 2, 167-186.

Schmidt, R.A . A schema theory of discrete motor skill learning. Psychol. Rev. *82*(1975)4,225-260.

Smith, E.E. Choice reaction time - an analysis of the major theoretical positions Psychol. Bull. *69*(1968)2,77-111

Stelmach, G.E. (Ed) Motor Control - Issues and Trends (Academic Press, New York 1976).

Van Galen, G P. Ambient versus Focal Information Processing and Single Channelness (Doct. Diss., Univ. of Nijmegen 1974).

Vince, M.A, The intermittency of control movements and the psychological refractory period, Brit J. Psych. *38*(1948)149-157

Welford, A T. Fundamentals of Skill (Methuen, London 1968)

Wickens, Ch D . A study of rapid controlled movement employing an isometric force sensitive control (unpubl. manusci , University of Michigan 1968)

Woltring, H J New possibilities for human motion studies by real-time light spot position measurement. Biotelemetry *1*(1974)3,132-146.

Woltring, H.J Calibration and measurement in 3-dimensional monitoring of human motion by optoelectronic means - I. Preliminaries and theoretical aspects Biotelemetry *2*(1975)3/4, 169-196, and - II Experimental results and discussion Biotelemetry *3*(1976) in print.

Woodworth, R.S The accuracy of voluntary movement Psychol Rev., Monogr Suppl *3*(1899)nr. 2. An excerpt is contained in D. Legge (Ed)· Skills (Penguin, Harmondsworth 1970)

Every little movement has a meaning all of its own.

Harbach and Hoschna, Song hit in "Madame Sherry" (1909).

3.2

Some Kinematic Aspects of the Control of Posture and Locomotion

Summary

As an illustration of the 3-dimensional reconstruction methodology of Chapter 2, the relations between the movements of a number of selected landmarks on a subject's body during locomotion have been investigated. It was found that arm and leg movements with respect to the trunk in the sagittal plane were highly correlated, both for single-legged and for biped locomotion, and that such correlations were small in the lateral direction. In single-legged hopping, the subject (who had suffered a right-knee patellectomy a few years ago) exhibited differential effects between the knees. The results are discussed in terms of the mutual dependence of posture preservation and locomotion.

Introduction

The quantitative analysis of the regulation between posture preservation and locomotion in man would appear to originate from Braune and Fischer's work (1895-1904) on human gait; as discussed in Chapter 4.1, they performed extensive studies on the path of the corporal centre of gravity based on analytical reconstruction in 3-D from photographed movement trajectories and on a mechanical approximation of human mass distribution. Indeed, the use of anthropometric data or of kinetic quantities is a prerequisite for such studies; however, some ideas on the coordination of limb and trunk movements might well be gleaned from purely kinematic research. Extensive work in this area has been conducted by Eberhart *et al.* (1947), especially in view of the design

of artificial limbs. More recently, Murray and associates (see Murray, 1967, for a review) reported various studies on both normal and pathological subjects; they concluded that sagittal displacements of trunk and limb in ordinary free-cadence walking were strikingly similar within and between subjects, but that transverse rotations of pelvis and thorax varied widely, suggesting that these actions may not be obligatory elements in normal gait. In addition, they observed a close temporal relationship between upper and lower limb movements in the sagittal plane.[1)]

It would seem that such relationships have usually been judged by visual inspection only, even though the parent data may be available through digitized film data or through electrogoniometry (see Adrian, 1973, for a technical review). Especially in the case of noisy or complex trajectory shapes, such judgments tend to become unreliable, and more analytical methods are called for. In the case of repetitive movements as typically occur in locomotion, a suitable approach is to employ correlation functions; the present chapter contains some preliminary results on the correlation between movement components in three orthogonal directions based on the methodological results of the Chapters 1 and 2.

Since the observations had to be made in a rather small room (depth 6 m), only a few locomotive strides could be studied per trial. Still, some striking correspondences between limb movements could be found.

Equipment and Task

The camera configuration corresponded to the one used for the observations on the hopping subject of Chapter 2.1.2 (see Fig. 3 on p. 90). The subject was equipped with 10 IR LEDs at the feet, knees, navel, hands, elbows, and on the head. Even though the marker positions did not correspond with body segment axes, the dynamic display of the reconstructed landmark positions as projected into the coronal and sagittal planes presented a realistic impression of human movement; only the horizontal view was usually rather unintelligible.

A single subject was employed, and he hopped toward the cameras (positive Z-direction) either on one leg or on both. Hopping movements were chosen for two reasons; first, it was anticipated that such movements may require additional balancing by upper body segments during the suspension phase as compared to ordinary free-cadence walking; secondly, this particular

[1)] From Atwater (1973), p. 219.

subject suffered a partial patellectomy in his right knee some years ago, and it was expected that single-legged hopping would reveal differential effects between the knees.

Data Processing

The raw camera observations were sampled at 15 msec intervals for 3 sec by a PDP-11/45 computer in an adjacent room, and they were further processed for 3-D reconstruction after movement trials. Subsequently, autocorrelation functions and crosscorrelation functions of selected landmark movements *with respect to* the navel landmark were assessed by Fast Fourier Transform (FFT) techniques (Oppenheim and Schafer, 1975; Rabiner and Gold, 1975); the latter landmark is assumed to represent a suitable estimate for the corporal centre of gravity. Since the central coordinate system of the subject's body was parallel to the object-space frame of reference (lateral - X; vertical - Y; frontal or locomotive - Z), it made sense to correlate the orthogonal movement components as reconstructed.

The filtered movement differences are rather noisy as is apparent from the various figures below; this is to be expected in view of their peak-to-peak amplitudes AMP which are in the order of 5 cm; the observation noise standard deviation (see Table III in Chapter 2.1.2, p. 86) at, for instance, Z = -0.5 m, is in the order of 2 mm, and the standard deviations of the filtered position estimate have been found to be 1.6 mm[2)] for the steady-state Kalman filter case. However, this high-frequency noise is smoothed in the various correlation functions which were assessed over a time shift range equal to the record length.

Experimental Results

The data shown in the Figs. 1 and 2 are based on the biped hopping experiment which was already discussed in Chapter 2.1.2 (see Fig. 5 on p. 91).[3)] It appears that ipsilateral and contralateral hand and knee movements in the sagittal plane are highly correlated as exemplified for the ipsilateral case in these figures. The low-frequency trend in the vertical Y-component dis-

[2)] For the Kalman filter parameters chosen (N = 3, T = 15 msec, SDX = 0.1 mm, SDZ = 5 mm; cf. Ch. 2.1.2), the *a posteriori* steady-state position standard deviations equal 3.7 mm; for SDZ = 2 mm and proportionally decreased SDX, the value 1.6 mm is obtained.

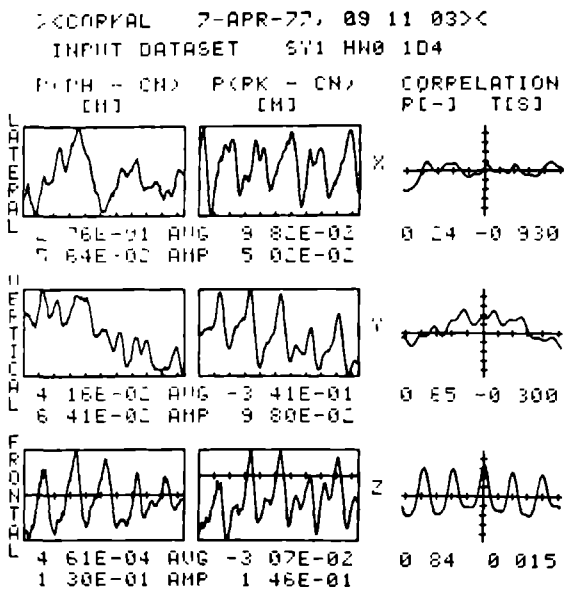


Fig. 1. Displacements of right hand and right knee with respect to the navel and crosscorrelation functions in the biped hopping experiment of Fig. 5 in Chapter 2.1.2 (p. 91).

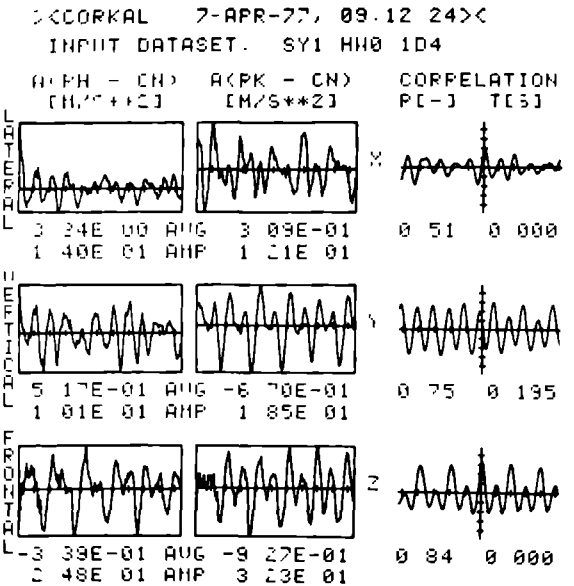


Fig. 2. Accelerations and crosscorrelations as in Fig. 1.

appears for the Kalman derivative estimates as apparent from Fig. 2. Especially the narrow Z-correlation peaks at $T_{shift} \cong 0$ reveal a high timing accuracy between hand (arm) and knee (leg) movements. For example, the crosscorrelation function of the Z-acceleration peaks at $r = 0.84$ for $T_{shift} = +3$ msec (as obtained by interpolating between the 15 msec sample values), and it drops to $r = 0.74$ for $T_{shift} = -9$ msec and $T_{shift} = +27$ msec. Since the forward movement of the knee entails an imbalance of the centre of gravity, the synchronous hand (arm) movement may be tentatively interpreted as a compensatory activity: both landmarks cross the navel in the Z-direction.

In the lateral direction, the crosscorrelation functions are much smaller in amplitude; similarly, the crosscorrelation functions of head and hands with respect to the navel have been found to be small, suggesting the operation of other balancing mechanisms in the coronal plane. introspectively, it was felt that force modulation in the active leg or legs was involved.

In a few cases, slower trends over strides correlated highly, suggesting a slow posture stabilization process superimposed on the basic hopping pattern.

As shown in Figs. 3-6, a marked difference was observed in the autocorrelation functions of the knee displacement with respect to the navel for left and right legged hopping (normal and operated knee, respectively) whereas the frontal and vertical components are synchronous over strides for either leg, this is not the case for the lateral component in left-legged hopping. Interestingly, it is the *operated* leg that exhibits synchrony in all directions (Figs. 5 and 6), whereas the normal leg reveals stride-independent movements in the lateral direction. It would seem that sagittal balancing is synchronous with locomotion in all cases, and that normal balancing in the coronal plane is independently controlled. The abnormal knee does not allow such control, and some form of balancing would seem to occur between strides. This appears also from the filtered Z-displacement curves in the Figs. 4 and 6: the jerky, positive values for the right knee correspond with extended knee flexion during ground contact, as opposed to the longer lasting knee extension in Fig. 4. Again from an introspective point of view, some time was spent in lateral adjustment during the stance phase.

[3] In these and other figures, the time increment per division is 0.25 sec. AVG denotes the mean signal value in m, m/s and m/s^2 for displacement P, velocity V, and acceleration A, respectively, and AMP denotes the signal peak-to-peak amplitude If the zero ordinate value fits within the signal frame, it is presented as a horizontal line. Furthermore, the correlation maxima R are given, together with their time shifts T, of course, for autocorrelation functions these values are 1 and 0, respectively. For the corporal landmarks, the following abbreviations are employed RH - right hand, LK and RK - left and right knee, respectively, CN - navel

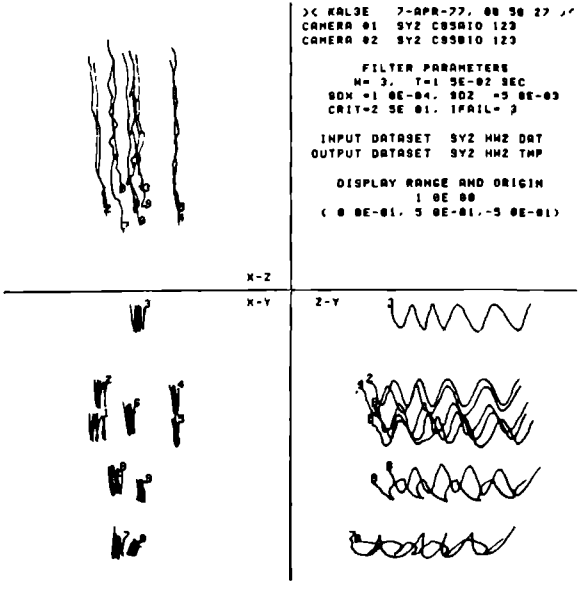


Fig. 3. Reconstructed trajectories for hopping on the left (normal) leg.

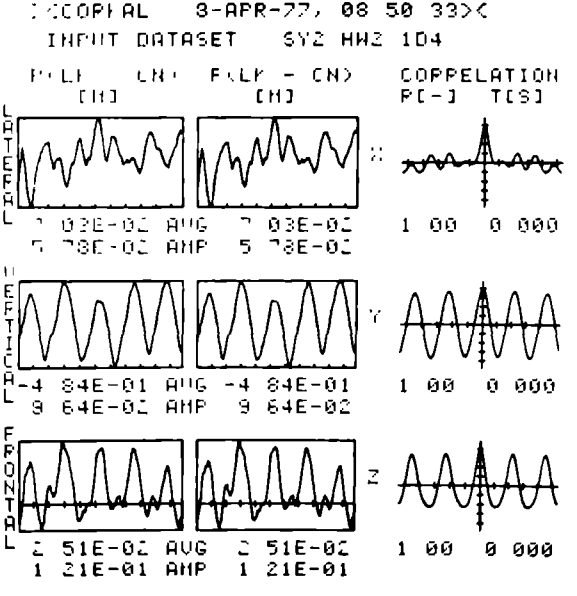


Fig. 4. Displacement of the supporting, left knee with respect to the navel and autocorrelation functions.

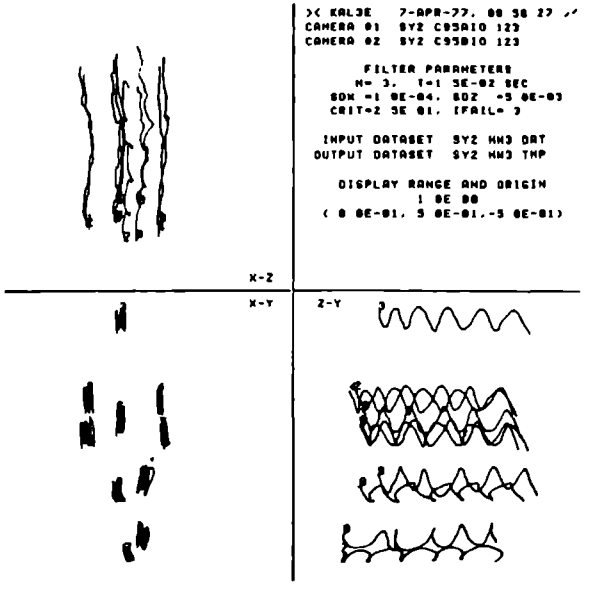


Fig. 5. Reconstructed trajectories for hopping on the right (operated) leg.

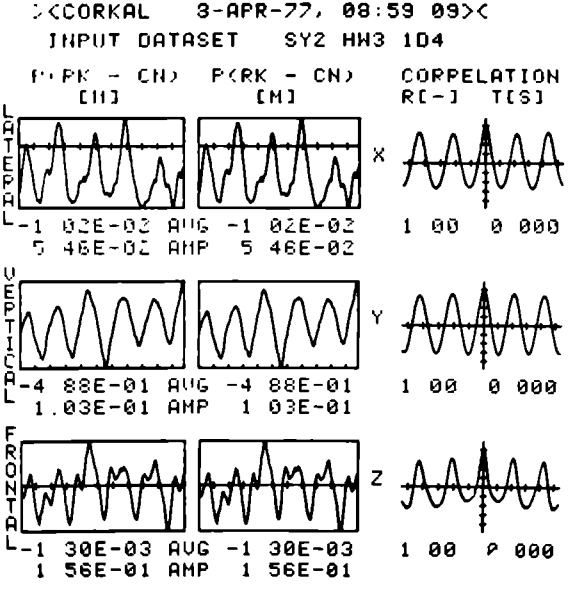


Fig. 6. Displacement of the supporting, right knee with respect to the navel and autocorrelation functions.

Discussion

The observations on a single subject, based on a few trial movements allow only tentative conclusions; yet, they demonstrate that conclusions on movement coordination may be obtained with little effort once one is able to assess 3-dimensional movement patterns: calculation of correlation functions is a comparatively straightforward procedure, and the computation time where FFT-methods are employed is in the order of a few seconds.

The kinematic data above suggest that the arms are involved in posture preservation in the sagittal plane. However, it is not clear to what extent active and reactive forces are responsible for this effect, and a kinetic analysis (see, e.g., Cappozzo *et al.* 1975; McGhee *et al.* 1976, and the references in Chapter 4.1) would be required to calculate muscle forces in the shoulder and elbow joints. If these forces are highly synchronous, there is more reason to assume the involvement of motor programs with or without afferent feedback for simultaneous posture preservation and locomotion.

An other limitation of the kinematic approach in the study of the relation between posture and locomotion control above is that no internal mechanisms mediating the sensory aspects are identified; current physiological research centres on the involvement of the stretch-reflex, the inner ear equilibrium organs and on the visual system in sensing posture disturbances (see Gurfinkel and Shik, 1973, for a review). Thus experiments as those reported above might be repeated with closed eyes, and perhaps even with anaesthetized tendon organs or muscle spindles, in order to evaluate to what extent these different stabilization systems are involved.

References

Adrian, M.: Cinematographic, electromyographic, and electrogoniometric techniques for analysing human movements. In: J.H. Wilmore (Ed.): Excercise and Sport Sciences Review *1*(1973).

Atwater, A.E.: Cinematographic analysis of human movement. *Ibid.*

Cappozzo, A. *et al.*: A general computing method for the analysis of human locomotion. J. Biomech. *8*(1975) 307-320.

Eberhart, H.D. *et al.*: Fundamental studies of human locomotion and other information relating to the design of artificial limbs. A Report to the National Research Council, Committee of Artificial Limbs (Univ. of California, Berkeley 1947).

Gurfinkel, G.S. and Shik, M.L.: The control of posture and locomotion. In: A.A. Gydikov *et al.* (Eds.): Motor Control (Plenum, New York 1973).

McGhee, R.B. *et al.*: Automatic estimation of joint forces and moments in human locomotion from television data. IFAC Symposium on Identification. Tblisi, U.S.S.R. 1976 (in print).

Murray, M.P.: Gait as a total pattern of movement; including a bibliography on gait. Amer. J. Phys. Med. *46*(1967) 290-333.

Oppenheim, A.V. and Schafer, R.W.: Digital Signal Processing (Prentice-Hall, Englewood Cliffs N.J. 1975).

Rabiner, L.R. and Gold, B.: Theory and Application of Digital Signal Processing (Prentice-Hall, Englewood Cliffs N.J. 1975).

Chapter IV

Developments in Movement Research and Summary

We will not anticipate the past; so mind, young people - our retrospection will now be all to the future.

Sheridan, The Rivals, Act. iv, sc. 2.

4.1

Marey Revisited: Prospect and Retrospect

Summary

In the present chapter, a somewhat eclectic survey is presented on developments and current trends in non-contacting kinesiological methods for motion registration, and in the investigation of human movement in psychology and related disciplines.

Measurement of Continuous Movement: Technology

Human and animal movements have fascinated many an investigator at least since Greek times, and attempts to observe and measure objectively have been manifold. Initially, nothing but careful observation (*motoscopy*) could be used as a research tool, leading, for instance, Aristotle to conclude that "positional change of the body during gait is caused by pushing off the body against the ground"[1], - a somewhat trivial conclusion at the present state of knowledge. Later work, as discussed by the Weber brothers (1836), includes that of Fabricius ab Acquapendente (1618) and of Galileo's pupil Borelli (1685); of course, Leonardo da Vinci's work on human anatomy is a landmark in its own right, and he investigated among others the movement of the body's centre of gravity during gait.[1]

[1] Quoted from Van Hussen (1973).

In 1836, the brothers Wilhelm and Eduard Weber published their book "Mechanik der menschlichen Gehwerkzeuge" (mechanics of human gait equipment) which is often conceded to be the first work of scientifically acceptable standard. In reviewing the previous litterature, they noted that a lack of experimental attitude and of proper instrumentation had hindered accurate and reliable observation of corporal movement during locomotion: "Da aber die Bewegungen des Körpers beim Gehen und Laufen sehr schnell vorübergehen und mehrere gleichzeitig geschehen, so ist eine genaue und zuverlässige Beobachtung dieser Bewegungen nur unter besonderen, dazu vorbereitenden Verhältnissen und durch die Anwendung von Instrumenten möglich, durch welche die Erscheinungen einzeln und von einander geschieden beobachtet werden können" (p. 384).

They noted that only Borelli (1685) and Gassendi (1592-1655) reported actual experiments in relation to human gait. Both were interested in the rectilinearity of human locomotion. Borelli erected two vertical poles, one behind the other, and found that he could not keep them in line when walking towards them; Gassendi simply observed the distance between hand and wall, with outstretched arm during walking (p. 384-385).

The Weber brothers had at their disposal only the limited equipment of their time - measuring lines and diopters (Bernstein, 1967), and clockworks. Yet they investigated succesfully for the first time the stance and swing phases during gait, and the relation between step duration and step length. Even though their *pendulum theory* of the swing movement was refuted in later research (see, e.g., Van Hussen, 1973), it would appear to be the first attempt at mathematically modelling aspects of human movement.

Carlet (1872) applied pneumatic methods to automate the registration of the step sequence: rubber balloons with two separate compartments were mounted under the shoes, and air pressure changes in these chambers were transmitted through rubber tubes for recording on a *kymograph* - a slowly rotating drum covered with smoked paper.

However, continuous registration of single- or multipoint movement trajectories was not possible with such and similar methods. The German physiologist Karl Hermann Vierordt (1881) described a charming approach employing thin ink jets which were sprayed from the body into vertical and horizontal directions; these jets hit long stripes of paper lying on the ground and hung alongside the subject's walking path. In addition, small cotton wads soaked in ink and attached to the shoes marked the shoe positions during locomotion; see Figs. 1 and 2.

In Paris, the physiologist Étienne Jules Marey started his investigations in animal and human locomotion around 1870, thus generalizing his previous work on muscle preparations as described in his lecture series "Le Mouvement

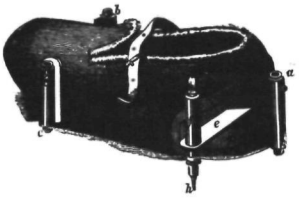


Fig. 1. Shoe used in Vierordt's experiment for continuous movement registration. a, b, c: cotton wads for shoe position marking; h: ink-spraying outlet. From Vierordt (1881), p. 7.

Fig. 2. Ink-spraying outlets for registration of vertical movement components, to be affixed to body extremities. From Vierordt (1881), p. 12.

dans les Fonctions de la Vie" (movement in the life functions). Marey was one of the most outstanding physiologists of his time, and he published extensively on the blood circulation, muscular contraction, respiration, human and animal locomotion, and on the flight of birds and insects (for a bibliography, see Anonymus, 1904); his book "La Méthode Graphique dans les Sciences Expérimentales" (1872, 1885) is a landmark on experimental methodology and instrumentation.

Originally relying on kymographic registration, he soon reverted to photography once sufficiently fast photographic material became available. In his book "Le Mouvement" (1894) he showed some of the trajectories obtained by photographing black subjects with white markers on the body. Employing stereophotography, even 3-dimensional observations were made, which could be instrumentally reconstructed by means of a stereoscope; he noted that an observer could even learn to fuse the stereoscopic photographs without the help of such a device.

In the United States, Muybridge had started his famous work on horse locomotion (Willmann, 1882; Haas, 1976) at about the same time. Using a series of cameras triggered one after another, he obtained details of horse locomotion hitherto unknown. In his earliest experiments, Muybridge made his horses trigger the cameras themselves, by the use of trip wires in the animal's path. At Marey's request, Muybridge attempted to make some observations on bird flight, and in 1881 he brought a number of photographs to Paris. However, as Marey (1885) relates, the timing aspects of these observations were unsatisfactory since the bird could not trigger photographic observations, as could Mybridge's horses. Therefore, Marey devised his "photographic gun" (see Fig. 3) which might be viewed as one predecessor of cinema-

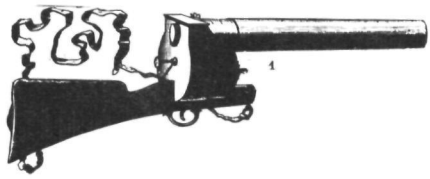


Fig. 3. Marey's photographic gun, allowing 12 photographs to be shot within 1 sec. From Marey (1885), App., p. 13.

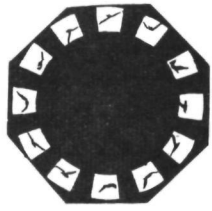


Fig. 4. A series of photographs on a flying gull, as obtained by means of the photographic gun. From Marey (1885), App., p. 15.

tography. By means of this device he could shoot 12 photographs within a second, each at 1/720 sec exposure time, without needing the animal for proper triggering of an exposure.

The mechanical problems incurred when he attempted to increase the exposure rate or number of equidistant exposures motivated Marey to develop his *chronophotographic* method. Basically, a single photograph is intermittently exposed to a white-clad subject moving against a blackened background, by means of a rotating disk with one or more apertures in front of the camera lens. In this fashion, Marey was able to register successive phases in walking, running, jumping, etc. as illustrated in Fig. 5. More abstract movement trajec-

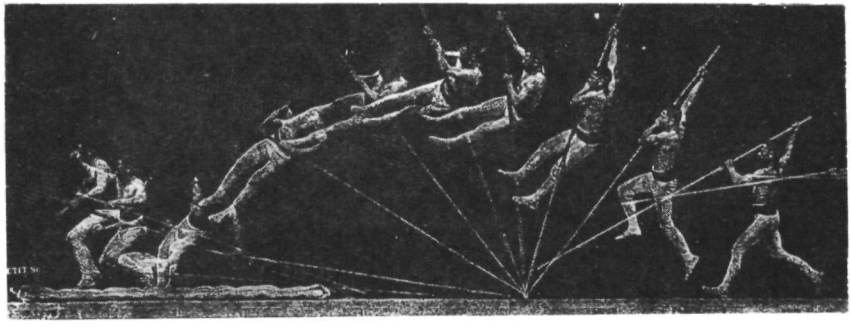


Fig. 5. Chronophotograph of pole-vaulting. From Marey (1894), p. 134.

tories were recorded by using reflective stripes or dots on dark subjects, as shown in two fairly famous photographs in Figs. 6 and 7.

Fig. 6. Black-clad subject with reflective markers on the body for chronophotographic movement observation. From Marey (1894), p. 60.

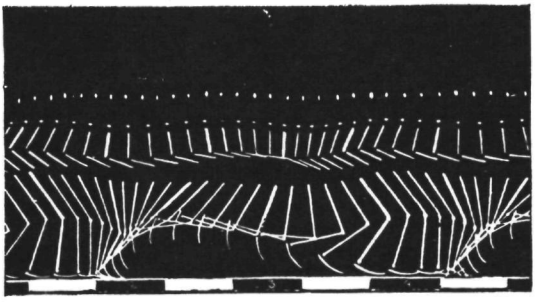


Fig. 7. Chronophotograph or *stick-diagram* of a runner. From Marey (1885), App., p. 34.

Marey had been forced to perform his experiments in broad sunlight as to obtain sufficient light for his photographs. His successors replaced the reflective tapes by incandescent light bulbs or flashing tubes. This alteration in technique was termed *cyclography* and developed to a high level at the Moscow school of Biometrics (Bernstein, 1967).

With typical precision and thoroughness, the German physiologist Otto Fischer, initially with his teacher Wilhelm Braune, performed extensive gait analysis on humans (1895-1904). In order to increase light intensity, they employed flashing tubes (Geisschler'sche Rohren) controlled by a Rhumkorff (spark) inductor: see Fig. 8. They were the first to report accurate, analytical reconstruction of 3-D trajectories from centrally projected observations (see Fig. 9), using 4 cameras: 2 observing the subject laterally, and 2 obliquely in front. It is impressive to realize how they ultimately assessed trajectories of corporal centres of gravity during locomotion as projected into the three coordinate planes (Fig. 10), by means of calculations extensively tabulated in their work which required many months to perform. In view of the time required to prepare an experiment (6-8 hours), it is not astonishing that they supported Vierordt's contention that the ink-spray method discussed above was more practical than photography from a clinical point of view, despite the larger errors made with this nonphotographic method.

In Russia, Bernstein and his coworkers perfected the cyclographic method, e.g. by slowly moving a film through a camera in the case of repeated movements (*kymocyclography*). For 3-D movement analysis, Bernstein (1930)

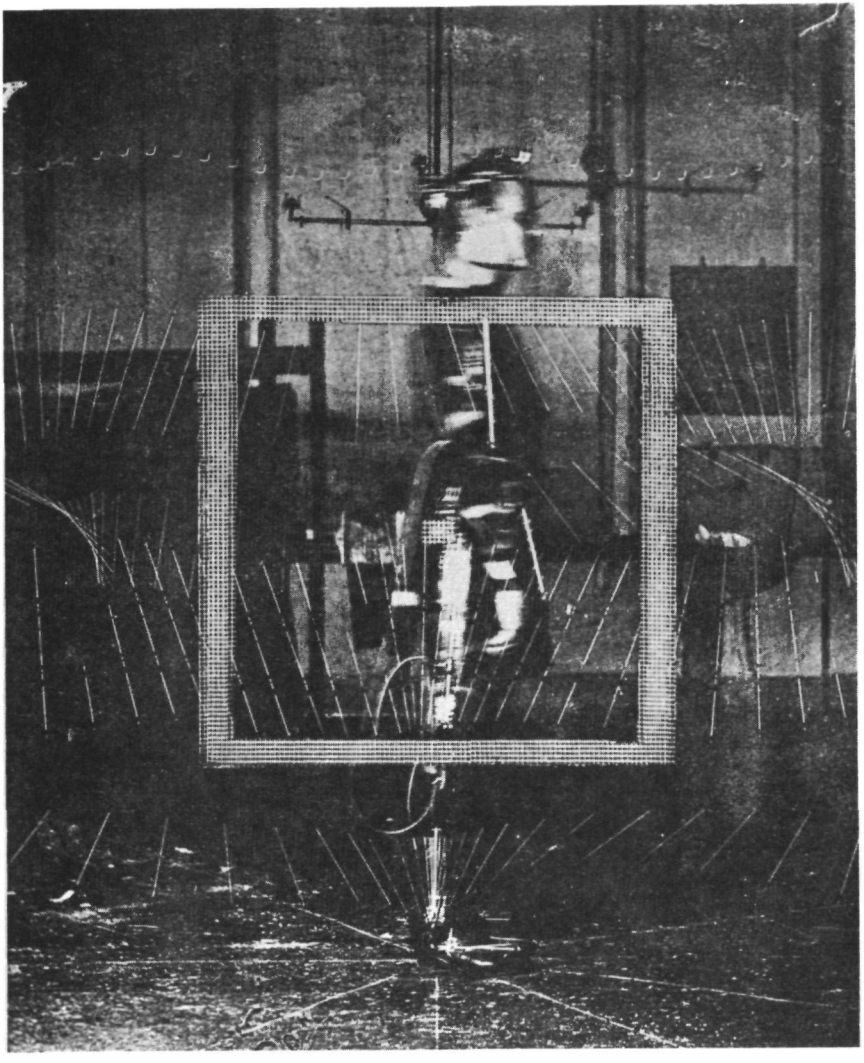


Fig. 8. Cyclogram or *stick-diagram* superimposed over a subject equipped with flashing tubes. The rectangle in the centre of the picture was added to facilitate image coordinate measurements. From Braune & Fischer (1895).

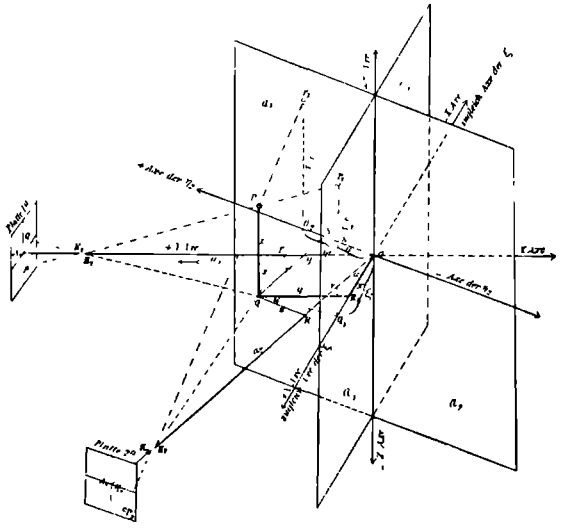


Fig. 9. Relation between object-space coordinates, camera lay-out and image coordinates. From Braune and Fischer (1895), p. 196.

rejected standard stereoscopic cameras because of their low depth accuracy (camera base 6.5 cm), and resorted to the use of *mirror kymocyclography*. Here, a mirror is placed at, e.g., 45° to the optical axis of a single camera, and each photograph or film contains a front view and side view of the movement pattern under study. Through this method, he could report spatial reconstruction errors below 1 mm per coordinate axis, for the investigation of repetitive skills such as punching.

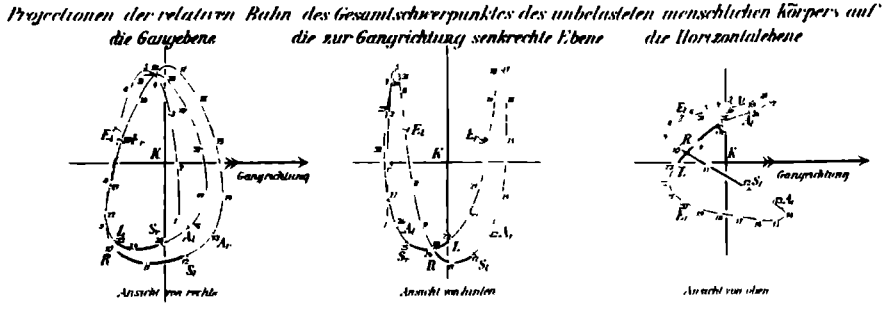


Fig. 10. Projections of the relative trajectory of the corporal centre of gravity for an unloaded subject. Left: side view; middle: rear view; right: top view. From Fischer (1899).

More recently, various other methods for noncontacting movement registration have been reported, and some of these were discussed in Chapter 1.2. In addition, mention should be made of the techniques developed by Eberhart & Inmann (1951), and of the UNOPAR system developed by Nadler & Goldman (1958). Here, an ultrasound source affixed to the body emits continuous sound waves which are captured by a number of remotely positioned microphones. Because of the Doppler effect, the velocity components of the sound source may be assessed by measuring frequency variations in the received signals, allowing the calculation of positions and accelerations through integration and differentiation, respectively.

An original method for multipoint movement observation was reported recently by Gueth *et al.* (1973) and Heinrichs (1974). Small photodiodes are mounted as landmarks on the subject's body, and a V-shaped light image is projected onto a rotating mirror. Thus, the light image is periodically swept around, and the time instants at which the legs of the V cause a photocurrent are linear functions of the horizontal and vertical directions of these diodes with respect to the mirror. At 8 m distance, a resolution between 0.7 and 1.4 mm was reported, for a scanning frequency of 40 Hz.

As indicated in previous chapters, these contemporary developments avoid the tedious and errorprone datareduction procedures of manual or semiautomatic film digitizing, albeit in some cases at the cost of wired body markers or insufficient spatial resolution. In a similar way, further data processing such as 3-D reconstruction, noise reduction and derivative assessment has been shown to be feasible through automated computation. In a sense, the explosive development in this field has enabled major parts of "lost labour" to turn from repetitive manual calculations *within* and *after* experiments to repetitive software developments *across* and *before* experiments. At the present time, hardware and software developments in biomechanics and kinesiology still proceed in a fairly independent fashion. Yet, for such matters as discussed in previous chapters including multitarget identification, 3-D reconstruction and derivative assessment, standard *motometric* equipment ought to be designed in a coordinated manner; current trends in microcomputer technology and real-time signal processing would seem to be most promising for this purpose. This would enable to relegate experiment-dependent data-processing to computer facilities at a higher level, e.g. conversion of *kinematic* quantities (positions, velocities, accelerations) to *kinetic* quantities (forces, moments, energies). Such calculations, moreover, are too complicated to be solved in today's relatively simple head-on computers.

Observation of Continuous Movement: Applications in the Life Sciences[1)]

The developments related above usually occurred in the context of applications to study the shape or *control* of human and animal movement. Initial work was physiologically or biomechanically oriented, some of it even raising lively interest outside the scientific community. Thus Muybridge's photographs on horse locomotion caused a sensation in art since it had not been realized previously how fast, biological movement really proceeds. With the advantage of hindsight, it is surprising to note that man's capacity to perceive at short exposure times (lightning!) had not been previously exploited for proper drawing, painting or sculpting of animal locomotion: it was only *after* photography had been developed substantially, that Galton (1882) described a device which allowed an observer to view millisecond exposures of fast phenomena.

Vierordt (1881) was much concerned with clinical work, and he compared extensively the ink-spray registrations obtained from normal and pathologic subjects including amputees, tabes dorsalis, spastics and haemiphlegia. Marey was more interested in what might now be called comparative kinesiology, although he did discuss various physiological applications of his chronophotographic method including observation of the heart's pumping action and of blood movement (Marey, 1894).

Interest in the *psychological* aspects of movement control seems to have started around the turn of the century. In his dissertation on learning and automation of movement in type-writing, Van der Veldt (1928) quoted the well-known studies of Bryan and Harter (1899) on acquisition of telegraphic skills, and those of Swift on learning stenography (1903) and typewriting (1904). Balancing between physiology and psychology, Woodworth (1899) published his paper on the control of fast, voluntary arm movements which has led to extensive research in intermittent movement control since Craik's (1947a, 1947b) and Vince's (1948a, 1948b) studies (see also Woodworth, 1903).

Isserlin (1914) started his paper "Ueber den Ablauf einfacher willkürlicher Bewegungen" (on the course of simple voluntary movements) in the style of his days: "Die Untersuchungen über welche hiermit berichtet wird, sollen einen Beitrag liefern zu den Erforschung der Willensvorgänge"[2)]; he

[1)] The author is indebted to a number of people both at Nijmegen University and elsewhere for discussions on the utility of objective motor measurements in Psychology and related disciplines.

[2)] The investigations reported here will contribute to the research into processes of conation.

proposed that research into the relation between "Einstellung" (set) and movement had unjustifiably been confined to simple measures such as reaction times, and that the *shape* of the movement expression had been neglected. Thus, one encounters the same situation as half a century before in the transition from the simple methods employed by the Weber brothers and Carlet to the more advanced and detailed registration procedures developed by Vierordt and Marey. Noting the bizarre movement patterns often observed in psychopathological cases, Isserlin proposed that differences in motor behaviour reflect underlying differences in man, and he set out to contribute to this issue by kymographic studies on the microstructure of simple, unidimensional, voluntary finger movements, differentiating between single, continued and repeated movements, and slow, continuously controlled movements.

Motor Disorders

The task of tapping between alternate targets has for a long time been used in the field of differential psychology, and Isserlin quoted a number of sources: "So hat Stern (1900) die Tempoklopfmethode als ein wertvolles Hilfsmittel für die Untersuchung der geistlichen Frische empfohlen. Und Kowalevski (1904) hat die Methode des Taktklopfens und Takthüpfens zur Kennzeichnung von Gemütszuständen benützt"[3] (p. 165). In the Netherlands Godefroy (1917, 1921) described the use of interrupted light photography (cyclography) to investigate epileptic dystaxia, registering movement trajectories between a series of alternate targets. By letting his subjects slowly move downwards, he could create a sort of kymocyclographic registration as shown in Fig. 11. Godefroy demonstrated some remarkable differences between normal and pathological movement patterns, and defined quantitive measures of *topodystaxia* or spatial irregularity, and *chronodystaxia* or temporal irregularity in motor control. For instance, he could capture epileptic seizures as typified by momentary movement interrupts at some place between targets which were often followed by increased acceleration after recovery. In addition, Godefroy confirmed earlier research from his *coordination photograms* that epileptic patients mainly try to correct targeting errors at the *end* of the initial step: only then, correction follows slowly and with difficulty (p. 273). This result would appear to be pertinent to those in the previous chapter on prior and current target information in movement control: it would be interesting to investigate to what extent distance reproduction based on early, current feedback is related to other measures of psychopathological behaviour,

[3] Thus, Stern (1900) recommended the tapping method as a valuable tool for the investigation of mental freshness. And Kowalevski (1904) has used the tasks of tapping and alternate tapping for the identification of dispositions.

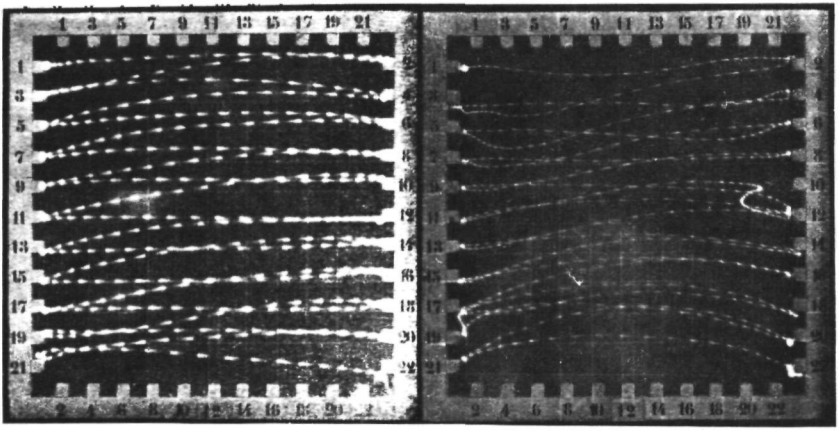


Fig. 11. Coordination photograms on a normal (left) and epileptic (right) subject. From Godefroy (1921), Figs. 3 and 20.

even more in view of Isserlin's (1914) distinction between "sensory" and "motor" biasedness in movement control which was also discussed in Chapter 3.1.

In many psychological applications the more complicated aspects of 3-dimensional movement reconstruction may be superfluous; yet, in the clinical field, subjects may be insufficiently malleable by instructions to allow clear observations from a single point of view. Especially in testing motor disorders, the introduction of *redundancy* in data acquisition, combined with proper data reduction may allow more objective diagnoses. Most tests in this field, e.g. the Oseretzky test for children (Oseretzki, 1931) and its later developments as, for instance, discussed by Stott (1966) suffer from lack of objectivity due to the *motoscopic* methods employed; one could imagine that the stick diagrams used in kinesiology and further formal data processing would allow procedural advantages (cf. Figs. 7 and 8).

As Wiegersma and Van der Velde (1975) have pointed out, a more stringent problem with such tests is that "the organism has many adaptive mechanisms at its disposal, through which certain motor tasks are performed in a seemingly normal manner, while in fact other than the usual means are employed in order to attain the intended goal. In addition, some children are masters at avoiding situations which might reveal their deficits" (p. 19). Therefore, objective measurements might enable the details of motor disorders to become apparent despite such deceiving behaviour.

Kiphard (1973; see also Hünnekens and Kiphard, 1963) has advocated the use of trampoline jumping to avoid such effects; the dynamic and statokinetic adaptation required in the *early* phase of the new situation litterally

throws the motorically disabled child out of balance, forcing it to abandon its compensatory strategies within the range of the most common motor activities. In this manner, a number of clearly recognizable abnormal motor reactions will become apparent.

Especially in place-restricted investigations such as the above, the use of *motometric* methodology including 3-D reconstruction of movement patterns might enhance the diagnostic value of test procedures, with the possibility to display motor patterns in terms of a frame of reference defined by the subject's own body (see, e.g., White, *et al.*, 1975). An interesting application was suggested some years ago by Hekking and Doombos in an unpublished report. Investigating the diagnostic value of trampoline jumping for organic cerebrality, they found a high correlation between Kiphard's test results, and more traditional measures including EEG, LPEG[4)], psychological tests and clinical-neurological diagnosis from "soft neurological signs". They concluded that the trampoline test was a useful device for deciding about organic cerebrality, and suggested further research to verify whether pshychological test + EEG + trampoline would be a useful indicator for LPEG, since the latter is a most painful and long-lasting procedure. Since their pilot studies were based on visual observation, simply dichotomizing the trampoline outcomes in a yes/no manner, one would expect that further research could profit from more detailed analysis.

Ergonomics or Human Factors

The investigation of normal motor behaviour in Psychology is mainly focused on motor *skills*. As Mulder *et al.* (1976) indicate, most psychonomic research in motor skills may be largely categorized into

a) skill acquisition, usually investigated in terms of SR-theory until Lashley's (1951) final reckoning with the limitations of this approach,
b) investigations of perceptual-motor tasks, usually control tasks in applied situations.

Prior to the era of industrial automation, not only control but also execution were the worker's tasks, and the classical Time and Motion studies as carried out by F.B. and L.M. Gilbreth (1911, 1917) highlight the application of movement studies of normal motor behaviour, deriving its inspiration from

[4)] The LPEG or *Lumbo-Pneumo-Encephalogram* is a Röntgenphotograph of the brain structures, where enhanced contrast is obtained by filling the cerebral ventricles with air. This is achieved by injecting air into the spinal column or occupit. The procedure entails much headache for the patient who must remain recumbent for some 2 days. Although present developments in tomography (brain scanners) do not require such drastic interference, it is not yet clear whether the LPEG is on its way to become obsolete. For furter details, see Kautzky and Zülich (1955).

F.W. Taylor's (1911) original work on "scientific management". As Taylor wrote: "Each job should be carefully subdivided into its elementary operations, and each of these ... should receive the most thorough time study"[5] in order to set standards for work quality and wage rates. First relying on only ordinary observation and stopwatch in determining form and time of movements, the Gilbreths successively employed cinematographic, cyclographic and stereocyclographic methods in their research of movements for establishing the "one best way of work" (Gilbreth and Gilbreth, 1917). In careful micromovement studies, they eliminated unnecessary movement elements and recombined the remaining ones into the fastest, least fatiguing, least errorprone, etc. form. In a time where industrial psychology was mainly oriented toward *selection* of workers for some given job, the work of Taylor and of his successors may be viewed as the first ergonomic investigation with clear attention to job analysis and design given the constraints of the human operator.

Notwithstanding the patent successes obtained with this approach, the strategy of decomposing complex movement patterns into standard elements which were recombined in some "optimal" sense has been criticized from various viewpoints. For one thing, in line with Gestalt tradition one may presume that "an organism is more than a sum of parts thrown together haphazardly" (Myers, 1923); also, it is doubtful whether individual differences are sufficiently small to be covered by some "one best way of work" based on observations of only a few subjects.

Further research in optimal work configurations and sequences for industrial tasks such as assembling, tool handling, and transportation was largely hindered by the tediousness of data analysis. Quoting from Welford (1974): "Scientific advances commonly come from a conjunction of interest in a problem and a technique for solving it. Interest in the high-speed motor skills in industry has been strong since the turn of the century, but the methods of recording fast movements in a continuous sequence were unsatisfactory, and the records were excessively laborious to analyse. The problems of recording and analysis were substantially solved in the early 1950's with the advent of machines which enabled events and times between events to be punched on paper tape, which could then be led into an electronic computer. However, by this time industrial interest had switched from high-speed manual work to the possible human problems involved in automation" (p. 390-391) - presumably, because of these very same developments. This argument obtains even more at the present day where paper tapes are becoming obsolete, and where movement trajectories of multiple corporal points may be directly fed into a computer, for slightly delayed or real-time data analysis.

[5] Quoted from Viteles (1932), p. 156.

The onset of automation in the military and industrial fields caused the second major surge of ergonomic task design, and this phase became nicknamed "knobs and dials ergonomics". Here, the study of movements was largely confined to SR-compatibility, and to simple dynamic properties including gain, friction and spring of controls (see, e.g., McCormick, 1970; Chapanis, 1965a). Complementary to this psychonomic, interface-oriented approach is the differential point of view pursued by Fleishman and his associates (Fleishman, 1967; for a review, see Fleishman, 1972), who aimed at developing a taxonomy of perceptual-motor tasks, trying to identify basic *abilities* (e.g., multilimb coordination, control precision, spatial visualization, rate estimation) which would serve as predictors for performance on actual, task-specific *skills*.

With the progress of automation, dominant motor requirements faded away and were replaced by calls upon more cognitive abilities; this third and current trend of *Man-Machine-System design* in Ergonomics views operator and machine as an organic whole, and centres upon task allocation between man and machine (see, e.g., Chapanis, 1965b): typically, monitoring tasks and decisions on general strategies in production, transport planning, and resource allocation are relegated to the human operator, and the details of such decisions are further elaborated by the machine. As a consequence, current performance requirements tend to focus on central properties such as intelligence and vigilance rather than on perceptual-motor abilities.

At the present time, it would seem that there is one field left where the investigation of human movement in all its complexity has not lost its importance: in vehicular control, the human operator still has to carry out a multitude of motor tasks which may require sufficiently fast action to warrant extensive studies of optimal movement patterns. This pertains especially to highly specialized spacecraft and aircraft, and to a lesser extent to automobile design. However, in the development of transport means for the handicapped, and of suitable prostheses and ortheses, there is a more earth-bound application for detailed studies of human movement patterns, a field to which also the Gilbreths dedicated themselves (1920).

In this context it is interesting to note that the photogrammetric principles discussed in Chapter 2 are currently being applied to prosthesis design (see, e.g., Herron and Karara, 1974); it would seem, however, that the psychological aspects in using artificial limbs are more of a motivational character than of a skill-learning nature. Still, in order to enable design of, for instance, suitable control algorithms for artificial limbs, there is a need for knowing more about "ordinary" movement patterns such as eating, drinking, manipulating, etc.: the design of artificial limbs which are both easy to use and easily accepted because of their "naturalness" has still a long way to go.

Sports and Physical Education

Rijsdorp (1971) has discussed the historical developments of what he calls "gymnology", or the science of physical education, sports, movement recreation and movement therapy; he noted that european interest throughout the ages has been highly correlated with times of social renewal: the classical Greek period, Renaissance, beginning and end of the 19th century, and the years following the two world wars - all periods with a tendency to renewal, a liberative character, an accent on humanity.

It would seem, however, that objective methodology in this field has been fairly limited until recently. Quoting again from Welford (1974): "The present wave of interest in motor performance seems to have been substantially fostered by the transformation of sport and athletics from healthy pleasure for participants to large-scale, lucrative entertainment and opportunities for political posturing, with consequent close attention to competitive standards. However much one may regret such a change, there is no doubt that it has produced some lively thinking by fresh minds on long neglected problems, and we may hope that with the aid of modern recording and analysing techniques, we shall soon see one of the last great unexplored areas of human performance adequately mapped" (p. 391). Current scientific interest in these aspects becomes clear from the emergence of new journals or annual volumes such as the Exercise and Sport Science Review, Journal of Motor Behaviour, and Journal of Human Movement Studies; the latter two are gradually taking over papers previously published in, for instance, Ergonomics and Human Factors, with more and more authors belonging to Departments of Kinesiology or Departments of Recreation and Physical Education.

An impressive amount of work on "Cinematographic analysis of human movement" was recently reviewed by Atwater (1973); it would appear, however, that most of these papers are based on relatively simple, 2-dimensional movement analysis from a single camera. In view of the 3-dimensional complexity of many typical skills in sports, e.g., in golfing, lawn-tennis, skiing, and swimming, research into the depth aspects of observed movement patterns is most necessary. Only since recently, some work is beginning to appear on full-scale 3-D analysis in sports (e.g., Borms *et al.*, 1973; Sodeyama *et al.* 1976), and especially on the level of computer simulation and interactive graphics (e.g., Miller, 1975; Riley & Garret, 1976).

Motor Skill Acquisition: Theoretical Issues in Psychonomy

Despite the multidimensional complexity of various industrial and sports tasks, experimental investigation of movement control in psychology, or *psychonomy*, has been largely confined to simple, unidimensional tasks. These involve usually only simple movements, for the analysis of internal mecha-

nisms which are quite complex even when investigated in only one dimension as, for instance, was the case in Chapter 3.1. Indeed, current theoretical topics in motor skills include anticipation and timing, feedback and central programming, speed and accuracy, acquisition and retention which may be easily investigated by means of simple, one-dimensional paradigms. On the other hand, in research concerning notions of single channelness or limited central processing capacity, systematic variation in the dimensionality of a motor task might enable to investigate changes from serial to parallel processing with skill acquisition.

An interesting approach in training complex motor patterns based on existing theories or skill acquisition would be to apply *biofeedback* principles. Thus, one might show on a display the movement trajectories of crucial landmarks on a subject's body during or after an attempt to perform, say, a golf swing, and superimpose normative trajectories (dotted, or in other colours) over these curves. Hopefully, such realtime movement observation and feedback will enhance learning, once suitable norms and feedback strategies can be established. A similar argument obtains for revalidation; in this manner, objective measurement of movement is not merely confined to diagnosis, but is extended to therapy. Biofeedback allows the presentation of both objective norms and objective error measures which are not simultaneously present in many other training procedures for dynamic skills.

The definition of proper norms remains a crucial question in this context, in a similar way as in the Gilbreths' Time and Motion Studies for establishing "the one best way of work": to what extent is, say, the backhand of today's world-champion in lawn-tennis normative for pupils if tomorrow's champion has a different shape or style?

Perception and Coding

If the study of *control* and *execution* of human movement is largely rooted in applied situations, the investigation of pure movement *perception* would appear to be a more academic issue. Classical work in this area has been performed by the Gestalt psychologists (see, for instance, Koffka, 1962), and by the Belgian Michotte who published in 1946 his work on "La Perception de la Causalité" in which he looked into the perception of mechanical actions by one body on an other. Later work includes that of the Swede Johansson (1973, 1976) who investigated the perception of human movement abstracted as a dot pattern on a display; these dots represent crucial landmarks on the body such as those in Fig. 5 of Chapter 3.1.2. Typically, variations are analysed in the subject's judgements when the movement pattern is experimentally changed.

An example of current work in this field is directed to the structure and use of motion verbs (Levelt *et al.*, 1976): "In spite of the fact that Michotte

was fully conscious of his experimental dependence upon the verbal reactions of his subjects, he never undertook a truly linguistical analysis of his subject's verbs of motion. That part of his work remained intuitive" (p. 1).

An old issue in perception has its more contemporary counterpart in motor control: the storage and novelty problem (cf. Chapter 3.1) for particular realizations of perceptual categories. Man and animal are able to recognize stimuli at orientations, directions and distances which they never have experienced before in exactly the same combination, and one would expect a more parsimonous way of stimulus decoding than the scanning of (infinitely) long lists of "templates", for instance, analysis in terms of "elementary features" (see, e.g., Bouma, 1976; Gibson, 1969; Sutherland, 1968), and of grammars which structure vocabularies of features (Leeuwenberg, 1969). Systematic investigation of human movement perception may profit from computer graphics based, for example, on viewing real 3-D movement patterns from different viewpoints, in order to find out how "general" internal pattern descriptions are: extensive experience with side and front views does not necessarily imply that the same object will be easily recognized from above.

An interesting topic would be to investigate whether any of the standard results on *categorical perception*, as discovered, for instance, in studies of stationary visual images and of (synthetic) speech perception (Liberman, *et al.*, 1967) apply also to the perception of motion (Marshall, 1977, pers. comm.). For example, is *between*-category discrimination easier (finer) than *within*-category discrimination? It would seem that a change of only one parameter in a moving dot pattern can cause the phenomenological appearance of the stimulus array to change from, for example, skating to walking or running. Again, one might systematically distort recordings of real movement patterns in 3-D as an experimental approach in such research.

An other classical issue, viz. the *nature-nurture* question in perception, a lively point of discussion for a number of centuries in philosophy, art, physiology, and psychology (Hochberg, 1962, 1971), might be investigated by similar means. Most contemporary infant studies on pattern perception have been confined to stationary stimuli in order to investigate depth perception and object recognition (see, e.g., Bower, 1964). In view of the "survival value" of movement perception, it would be interesting to find out to what extent recognition of certain movement patterns is innate; for example, one might investigate whether dot patterns of moving human subjects are given more attention than "nonsense" patterns with similar statistical properties, by means of eye/head movement monitoring and operant conditioning procedures (Marshall, 1977, pers. comm.). Innateness of essential perceptual categories might correspond to innateness of crucial movement patterns such as the gait-like movements observed in newly-born infants (see Conolly (1970)

on the nature-nurture question and the ensuing discussion on the possibility of artefacts in such gait patterns, and also Conolly & Bruner, 1974).

Social and Developmental Studies

Other applications where movement patterns cannot be confined to the simple, one-dimensional paradigms which are so often used for experimental work with normal adults include developmental studies, social interaction and nonverbal expressive behaviour. In such studies, the use of artificial landmarks - whether wired or not - may often be unacceptable, and this will necessitate human intervention in the analysis of raw film-data or TV-data However, spatial reconstruction and derivative assessment may still be achieved through automated data processing

Until recently, "the two fields of developmental studies and the analysis of adult skills have had curiously little influence on each other" (Kay, 1970, p. 140), despite shared interest in the structure of 'quanta of skilled performance', and in the changes in such modular units with progress of learning (Stelmach, 1976) and maturation (Conolly, 1970, Conolly and Bruner, 1974), respectively Typical experiments with children include tapping between alternate targets, manipulative tasks such as grasping, transportation and releasing, and more detailed visuo-motor control tasks such as inserting blocks of different shapes into corresponding holes Again, the "molecular" approach of recording and analysing movement trajectories may reveal subdecisions during movement and chaining of more elementary movements, this may tell something about changes in the hierarchy or *syntax* of movement control, and of the complexity and smoothness of elementary movement "chunks", in a manner reminiscent of Bryan and Harter's findings in telegraphy (1889).

As in the case of adult subjects, motor disorders in children may be objectively investigated along such lines For instance, Kalverboer (1975) investigated MBD (Minimal Brain Damaged) pre-school children by means of observational, ethological[6] methods, through video-recordings which were extremely laborious and errorprone to analyse. Schellekens (1976, pers comm) has proposed employing typical psychonomic motor skill concepts and movement trajectory measurements in order to compare normal and problem-children between 7 and 11 years of age, the latter usually attend so-called LOM-schools[7] and frequently display various MBD-symptoms

A further step in data processing beyond geometrical movement trajectories is to employ some suitable *movement notation system* in order to obtain a "choreography of movement display" (Golani, 1976) slightly reminis-

[6] For a discussion on the term ethology, see Beer (1975)

[7] LOM, or "Leer- en Opvoedings Moeilijkheden" learning and educational difficulties (Dutch)

cent of musical notation. Through such encoding procedures, complex movement patterns may be readily arranged for visual inspection. The "therbligs" devised by F.W. Gilbreth (1911) are an early example of such a system. A recent version is the Eshkol-Wachman movement notation system (Eshkol and Wachman, 1958) which has been applied to the description of a wide variety of subjects including classical ballet, physical education, folk dances, sign language of the deaf, and ethology (Golani, 1976). Through such notation systems, the investigation of verbal versus nonverbal expressive behaviour is greatly facilitated.

Plooy (1976) has suggested employing the Eshkol-Wachman notation system in order to investigate pre-verbal communication in infants. Since interaction between infants and adults is unquestionably 3-dimensional, the feasibility of applying photogrammetric principles for spatial reconstruction of movements from direct and mirrored TV-data is currently under study.

Concluding Remarks

In 1868, Marey wrote in his first lesson on "Le Mouvement dans les Fonctions de la Vie": "Pour bien juger de la direction qu'on doit suivre, il faut de temps à temps regarder en arrière, considérer l'espace que l'on a parcouru, se rappeler les détours, les dangers, les difficultés de la route. Ce détour sur le passé est la préparation la plus utile à une étape nouvelle; il nous permettra d'arriver à notre but plus vite et plus sûrement que nos prédécesseurs. C'est ainsi que l'expérience du passé profite à l'avenir, c'est pour cela que la marche vers le progrès s'accelère sans cesse, et qu'on voit de nos jours plus de découvertes se produire en dix ans qu'ils n'en apparaissaient autrefois dans un siècle"[8]. Even if 19th century optimism on the progress of science and technology is being replaced at present by an increasing awareness of possibly adverse effects on our environment, Marey's contentions of looking back on behalf of the future and of accelerated developments still hold. Thus, the "New Possibilities for Human Motion Studies..." of Chapter 1.2 will probably become outmoded in a few years' time, especially in view of current develop-

[8] For properly judging which direction one should follow, it is necessary to look back once in a while, to consider the space one has just passed through, to recollect the detours, the dangers, the difficulties encountered on the way. This digression into the past is the most suitable preparation for a new stage; it enables us to achieve our goal more quickly and more surely than our predecessors could. It is in this sense that the experience of the past is profitable for the future, that the march to progress accelerates unrelentingly and that in our time more discoveries have been made in 10 years than during a century before (Marey, 1868, p. 2).

ments in solid-state imagery. Nevertheless, "les appareils enregistreurs permettent maintenant d'aborder des recherches qui autrefois eussent été impossibles"[9], such as the use of real-time movement feedback in sports training and in revalidation which was suggested in the previous section of this chapter.

Research into movement coordination in terms of the motor *commands* issued by the moving subject requires conversion from kinematic to kinetic quantities as briefly indicated in Chapter 2.1.2 and in the first section of the present chapter. Even more than 3-D trajectory reconstruction, the complexity of the calculations involved in the solution of the equations of motion was prohibitive prior to the development of computer technology. The usual approach today is to view the moving subject as a *biokinematic chain* whose *links* have known masses and moments of inertia (see references 4, 7, 10, 15 in Chapter 2.1.2 on methods of assessing these anthropometric parameters), with frictionless joints connecting the links. The classical, analytical approach along Lagrangean or Hamiltonian lines (Whittaker, 1937; Chace, 1967) becomes unwieldingly complex with increasing number of links, and even more so when 3-D motion is analysed. Procedures which are more amenable to computer implementation include free-body methods (Khandelwal and Frank, 1974) and sequential link analysis (Vukobratovic and Stephanenko, 1973); for a review, see Paul (1974). Current approaches are often based on matrix methods, e.g. Hemani *et al.* (1975), Velikson and Chkhaidze (1975), and Kulakov *et al.* (1975); see also Aleshinsky and Zatziorsky (1975, 1976).

Even if such kinetic quantities are obtained, there is no direct correspondence with muscular commands, and one is successively confronted with modelling input-output relationships of muscles *in vivo* (e.g. Dijkstra, 1973), control input distribution in the currently known muscle input channels: α- and γ-efferent control (Granit, 1970, 1973; Matthews, 1972), and with feedback-loops on cortical and cerebral levels. It is almost trite to say that this 'moving into man' is the physiological approach, and hardly considered in psychological training procedures. Yet, research into the complex relationships between force patterns and movement trajectories in fast motor skills might perhaps improve training methodology. For instance, in fast skills there is hardly any time to use feedback during movement execution, and the subject must prepare a complete motor programme prior to movement execution. Thus rather then letting the subject solve the 'inverse problem of dynamics' of generating force patterns until trajectory criteria are met, one might present biofeedback on the actually generated force pattern, and let him compare this with a required pattern. At the present stage of computer technology, the technical problems for such procedures are still enormous, but one may hope

[9] The registration apparatuses now enable us to commence investigations which would have been impossible before (Marey, 1868, p. V).

that ultimately this "bypassing of the plant in a closed control loop" will be feasible.

In his review paper on "Recent advances in instrumentation and methodology of biomechanical studies", Cavanagh (1976) wrote: "It is vital that we realize that instrumentation and methodology are not ends in themselves. With equipment no more complicated than steel balls, planks of wood, and water clocks, Galilei performed experiments that were fundamental to our view of mechanics. In the face of a contribution that represented so much with so little experimental sophistication, it is a challenge to us to contribute a little with so much" (p. 409-410). Indeed, this proposition may be true from a research point of view, where lack of instrumentation and methodology may sometimes be traded for enthusiasm and many hours of 'lost labour'. However, for all practical purposes instruments must be simple and methods should be easy to use and to teach. Some of the issues raised in the present thesis are intended to contribute to this "end in itself"; at a later stage, they may "contribute to the research into processes of conation" as Isserlin would have it 63 years ago.

References

Aleshinsky, S.Yu. and Zatziorsky, V.M.: Modelling the spatial movement of man. Biophysics *20*(1975) 1144-1151. Also in: P.V. Komi (Ed.): Biomechanics V-B, p. 387-394 (Univ. Park Press, Baltimore 1976).

Anonymus: Catalogue des traveaux publiés par É.J. Marey. Arch. Ital. Biol. (Torino) *51*(1904) 489-498.

Atwater, A.E.: Cinematographic analysis of human movement. Exerc. Sp. Sc. Rev. *1*(1973) 217-258.

Beer, C.G.: Was Professor Lehrman an Ethologist? Anim. Behav. *23*(1975) 957-964.

Bernstein, N.: Analyse der Körperbewegungen und Stellungen im Raum mittels Spiegel-Stereoaufnahmen. Arbeitsphysiologie *6*(1930) 3.

Bernstein, N.: The Co-ordination and Regulation of Movements (Pergamon Press, Oxford 1967).

Borelli, G.A.: De Motu Animalium, I (Lyon 1680).

Borms, J.; Duquet, W. and Hebbelinck, M.: Biomechanical analysis of the full-twist backsomersault. In: S. Cerquiglini *et al.* (Eds.): Biomechanics III (Univ. Park Press, Baltimore 1973, and Karger, Basel 1973).

Bouma, H.: Perceptieve functies. In: J.A. Michon, E.G.J. Eijkman, L.F.W. de Klerk (Eds.): Handboek der Psychonomie, Ch. 8 (Van Loghum Slaterus, Deventer 1976).

Bower, T.G.R.: Discrimination of depth in premotor infants. Psychon. Sc. *1*(1964) 368.

Braune, C.W. and Fischer, O.: Der Gang des Menschen I. Abh. Math. Phys. Cl. Kön. Sächs. Ges. Wissensch. *21*(1895) 153-322.

Bryan, W.L. and Harter, N.: Studies on the telegraphic language. Psych. Rev. *6*(1899) 345-375.

Carlet, M.G. Sur la Locomotion humaine - étude de la marche. Ann. Sc. Nat. XVI.1. art. no. 6 (1872).

Cavanagh, P.R.. Recent advances in instrumentation and methodology of biomechanical studies. In P V. Komi (Ed.) Biomechanics V-B, p. 399-411 (Univ. Park Press, Baltimore 1976).

Chapanis, A. Man-Machine Engineering (Wadsworth, Belmont 1965a).

Chapanis, A. On the allocation of functions between men and machines. Occup. Psychol. *39*(1965b) 1-11.

Chase, M.A.: Analysis of the time-dependence of multi-freedom systems in relative coordinates. J. of Engng for Ind., Trans. ASME, ser. B, *19*(1967) 111.

Conolly, K. Skill development - problems and plans. In. K. Conolly (Ed.) Mechanisms of Motor Skill Development, p. 3-21 (Acad. Press, London 1970).

Conolly, K. and Bruner, J.S. The Growth of Competence (Acad. Press, London 1974).

Craik, K.J.W. Theory of the human operator in control systems. I - The operator as an engineering system. Brit. J. Psychol. *38*(1947a) 56-61, II - Man as an element in a control system Brit. J. Psychol. *38*(1947b) 142-148.

Dijkstra, S.J. Evaluation and Simulation of Fast Arm Movements. Doct. Diss., State University of Utrecht, Dept. of Physics (Utrecht 1973).

Eberhart, H.D. and Inmann, V.T An evaluation of experimental procedures used in a fundamental study of human locomotion. An. N.Y. Acad. Sc *51*(1951) 1213-1228.

Eshkol, N. and Wachman, A. Movement Notation (Weidenfeld and Nicholson, London 1958)

Fabricius ab Aquapendente, H. De Motu Locali Animalium Secundum Totum (Batavia 1618).

Fischer, O. Der Gang des Menschen. Abh Math. Phys. Cl. Kön. Sachs. Ges. Wissensch.:II - *25*(1899) 1-163, III - *26*(1900) 87-185; IV - *26*(1901) 471-596, V - *28*(1903) 321-425, and VI - *28*(1904) 533-623.

Fleishman, E.A. Performance assessment based on an emperically derived task taxonomy. Human Factors *9*(1967) 349-366.

Fleishman, E.A.. On the relation between abilities, learning, and human performance. Am. Psychol. *27*(1972) 1017-1032.

Galton, F.. Rapid-view instrument for momentary attitudes. Nature *26*(1882) 249-251.

Gassendi, P. (1592-1655) De Vi Motrice et Motionibus Animalium (P. Gassendi Opera, Vol. II, Book XI, Florence).

Gibson, E.J. Principles of Perceptual Learning and Development (Appleton-Century-Crofts, New York 1969).

Gilbreth, F.B. Motion Study (Van Nostrand, New York 1911).

Gilbreth, F.B. and Gilbreth, L.M. Applied Motion Study (McMillan, New York 1917).

Gilbreth, F.B. and Gilbreth, L.M. Motion Study for the Handicapped (Constable, London 1920).

Godefroy, J.C.L.. Light-line registration of movement. Psychiatr. en Neurol. Bladen *21*(1917) 431-439.

Godefroy, J.C.L. Dystaxia epileptica - an experimental psychological study regarding motory disturbances. Psychiatr. en Neurol. Bladen *25*(1921) 68-129, 272-275.

Golani, L.. Homeostatic motor processes in mammalian interactions. Ch. 2 in P.P.G. Bateson and P.H. Klopfer (Eds.). Perspectives in Ethology, Vol. 2 (Plenum, New York 1976).

Granit, R. The Basis of Motor Control (Acad. Press, London 1970).

Granit, R. Muscle sense, proprioception and the control of movement. In A.A. Gydikov, N.T. Tankov, and D.S. Kosarov (Eds.) Motor Control, p. 1-13 (Plenum, New York 1973).

Gueth, V. *et al.* Eine Methode zur Chronozyklographischen Bewegungsaufzeichnung mit einem Prozessrechner. Int. Z. angeew. Physiol *31*(1973) 151-162

Haas, R.B Muybridge - Man in Motion (Univ of Calif Press, Berkeley 1976)

Heinrichs, W. Eine digitale Zeitmesseinrichtung hoher Auflosung zur Weiterentwicklung der chronocyclographischen Bewegungsaufnahme mit Prozessrechner. Europ J. Appl. Physiol. *32*(1974) 227-238.

Hemani, H , Jaswa, V C., and McGhee, R.B. Some alternative formulations of manipulator dynamics for computer simulation studies Proc. 13th Allerton Conf. on Circuit and System Theory (Univ. of Illinois, 1975)

Herron, R E. and Karara, H.M. (Eds.) Biostereometrics 74 - Proc. of the Symp of Comm. V, ISP (Am. Soc. Photogrammetry, Falls Church 1974).

Hochberg, J Naturism and empiricism in perception. In L. Postman (Ed) Psychology in the Making, p. 255-330 (Knopf, New York 1962).

Hochberg, J. Perception - II. Space and movement. Ch 13 in J W. Kling and L A. Riggs (Eds.) Woodworth & Schlosberg's Experimental Psychology (Methuen, London 1971).

Hünnekens, H and Kiphard, E.J. Motoskopische Untersuchungen und Betrachtungen beim Trampolinspringen. Acta Paedopsychiatrica *30*(1963) 232-247, 324-341

Isserlin, M Ueber den Ablauf einfacher willkurlicher Bewegungen. Psychol. Arb. *6*(1914) 1-195.

Johansson, G Visual perception of biological motion and a model for its analysis. Perc and Psychoph. *14*(1973) 201-211.

Johansson, G. Spatio-temporal differentiation and integration. Psychol. Res *38*(1976) 379-393.

Kalverboer, A.F A Neurobehavioural Study in Pre-School Children (Heinemann, London 1975).

Kautzky, R. and Zulich, K.J Neurologisch-neurochirurgische Rontgendiagnostik und andere Methoden zur Erkennung intrakranialer Erkrankungen (Springer, Berlin 1955).

Kay, H. Analysing motor skill performance In K.J. Conolly (Ed) Mechanisms of Motor Skill Developments, p. 139-159 (Acad. Press, London 1970).

Khandelwal, B.M. and Frank, A.A. On the dynamics of an elastically coupled multi-body biped locomotion model. Proc. 1974 JACC (Austin, Texas 1974).

Kiphard, E J Bewegungs- und Koordinationsschwachen im Grundschulalter (Hoffman, Schorndorf bei Stuttgart 1973).

Koffka, K. Principles of Gestalt Psychology (Routhledge & Kegan Paul, London 1962, 5th ed.).

Kowalewski, P J Studien zur Psychologie des Pessimismus (Wiesbaden 1904).

Kulakov, F M., *et al.* Computer modelling of human movements. Biophysics *20*(1975) 1139-1144.

Lashley, K.S. The problem of serial order in behavior. In L.A. Jeffres (Ed) Cerebral Mechanisms in Behaviour, p. 122-130 (Wiley, New York 1951).

Leeuwenberg, E.L.L. Quantitative specification of information in sequential patterns Psychol. Rev. *76*(1969) 216-220

Levelt, W.J.M., Schreuder, R , and Hoenkamp, E C M Structure and use of verbs of motion. Cath. Univ. of Nijmegen, Lab of Psychol , Int Rept 76FU06 (Nijmegen

1976). Also to appear in Campbell and Smith (Eds.) Proc. of the Conf on the Psychology of Language (Stirling 1976).

Liberman, A.M., *et al.*. Perception of the speech-code. Psychol Rev. *74*(1967) 431-461.

Marey, É.J. Le Mouvement dans les Fonctions de la Vie (Baillière, Paris 1868).

Marey, É.J. La Méthode Graphique dans les Sciences Expérimentales (Masson, Paris 1872, 2nd ed. with supplement "Le Développement de la Méthode Graphique par la Photographie", 1885).

Marey, É.J. Le Mouvement (Masson, Paris 1894).

Matthews, P.B.C.. Mammalian Muscle Receptors and their Central Actions (Arnold, London 1972).

McCormick, E J. Human Factors Engineering (McGraw-Hill, New York 1970; 3rd ed).

Michotte, A La Perception de la Causalité (Inst. Sup. de Philos , Louvain 1946).

Miller, D.I. Computer simulation of human motion In D.W. Grieve *et al.* Techniques for the Analysis of Human Movement, Ch 3 (Lepus, London 1975).

Mulder, G., Michon, J.A., and Moraal, J. Motorische vaardigheden, Ch. 17 in J.A. Michon, E.G.J. Eijkman, and L F W. de Klerk Handboek der Psychonomie (Van Loghum Slaterus, Deventer 1976).

Myers, C.S. The efficiency engineer and the industrial psychologist. J Nat Inst. Ind Psychol. *1*(1923) 170.

Nadler, G. and Goldman, J.D. The UNOPAR. J. of Ind. Engng *9*(1958) 58-65

Paul, B. Analytical Dynamics of Mechanisms - a Computer Oriented Overview Univ. of Penn., Dept. of Mech. Engng & Appl. Mechanics, MEAM Rept 74-1 (Philadelphia 1974).

Oseretzky, N. Psychomotorik, Methoden zur Untersuchung der Motorik Beiheft, Z. fur angew. Psychol. (Leipzig) *57*(1931).

Plooij, F X. The development of pre-verbal communication in the mother-infant interaction - methodological aspects Cath. Univ. of Nijmegen, Lab of Psychol , Int. Rept. 76ON07 (Nijmegen 1977).

Riley, D.R. and Garrett, G.E. Dynamic interactive computer graphics package for human movement studies. In P.V. Komi (Ed) Biomechanics V-B, p 371-379 (Univ. Park Press, Baltimore 1976).

Rijsdorp, K. Gymnologie (Spectrum, Utrecht 1971).

Sodeyama, H *et al.* Study of the displacement of a skier's center of gravity during a ski turn. In P.V. Komi (Ed.) Biomechanics V-B, p. 271-276 (Univ Park Press, Baltimore 1976).

Stelmach, G.E. (Ed.) Motor Control - Issues and Trends (Acad. Press, New York 1976).

Stern, W. Psychologie der Individuellen Differenzen (Barth, Leipzig 1900).

Stott, E H. A general task of motor impairment for children. Dev. Med. & Child Neurol *8*(1966) 523-531.

Sutherland, N.S Outlines of a theory of visual pattern recognition in animals and man. Proc. Roy. Soc. B *171*(1968) 297-317.

Swift, E.J. Studies in the psychology and physiology of learning. Am. J. Psychol. *14*(1903) 201-251.

Swift, E.J. The acquisition of skill in typewriting. Psychol. Bull *1*(1904) 295-305.

Taylor, F.W. The Principle of Scientific Management (Harper, New York 1911).

Van der Veldt, J. l'Apprentissage du Mouvement et l'Automation (Inst. de Philos., Louvain 1928).

Van Hussen, F.A.J. Electrogoniometrie van de knie - een klinische kinesiologische studie (Walburg, Zutphen 1973).

Velikson, V.M. and Chkhaidze, L.V. Certain aspects of the mathematical modelling of the dynamics of biokinematic chains. Biophysics *20*(1975) 143-147.

Vierordt, K.H.: Das Gehen des Menschen in gesunden und kranken Zustanden nach selbstregistrirenden Methoden dargestellt (Laupp, Tubingen 1881).

Vince, M.A.: The intermittency of control movements and the psychological refractory period. Brit. J. Psychol. *38*(1948a) 149-157.

Vince, M.A.: Corrective movements in a pursuit task. Quart. J. Exp. Psychol. *1*(1948b) 85-103.

Viteles, M.S.· Industrial Psychology (Norton, New York 1932).

Vukobratovic, M. and Stephanenko, J.: Mathematical models of general anthropomorphic systems. Math. Biosciences *17*(1973) 191-242.

Weber, W. and Weber, E.. Mechanik der Menschlichen Gehwerkzeuge (Dieterich, Göttingen 1836).

Welford, A.T.: On the sequencing of action. Brain Res. *71*(1974) 381-392.

White, A.A. III, *et al.* A system for defining position and motion of the human body parts. Med. Biol. Engng *13*(1975) 261-265.

Whittaker, A.T.. A Treatise on the Analytical Dynamics of Particles and Rigid Bodies (Dover, New York 1937).

Wiegersma, P.H. and Van der Velde, A.: De Hamm Marburger Korperkoordinationstest fur Kinder (in Dutch) (Swets & Zeitlinger, Amsterdam 1975).

Willmann, - : *The Horse in Motion as Schown by Instantaneous Photography* (Turner, London 1882).

Woodworth, R.S. The accuracy of voluntary movement. Psychol. Rev., Monogr. Suppl. *3*(1899) nr. 2. An excerpt is contained in. D. Legge (Ed.). Skills (Penguin, Harmondsworth 1970).

Woodworth, R.S.: Le Mouvement (Doin, Paris 1903).

Scire est mensurare

Joh. Kepler

4.2

Summary

The present investigation is concerned with measurement and control of human movement, and may be subdivided into a technically and a behaviourally oriented part.

The investigation of the human motor system entails serious problems of data processing due to the large variability in possible movement patterns. Therefore, a system which incorporates fully automated data acquisition and data reduction would be preferable to more traditional approaches which are usually - at least in part - based on manual digitizing of photographic or cinematographic data.

A comparison of some optoelectronic measurement procedures resulted in an experimental system which exploited the *lateral photoeffect* for position determination of small light sources affixed to the body (1.2). In anticipation of an improved and commercial version which was independently developed at Chalmers' Institute of Technology, Gothenburg, Sweden, a theoretical investigation was conducted of the position-sensitive detector employed in both systems (1.3).

The improved commercial system enabled to measure 3-dimensional light point trajectories by means of two cameras, and a part of the present thesis has been devoted to the calibration and measurement aspects thereof. For this purpose, *photogrammetric* principles were employed in order to determine such system parameters as camera position, attitude, effective focal length and image distortion, by observing a set of calibration points which are distributed in object space in a suitably chosen fashion (Chapter 2).

In movement research there is often a need for simultaneous estimation of movement trajectories and time derivatives; the use of Kalman filtering methods was discussed (2.1.2) and illustrated with an example of 3-D gait reconstruction.

Since the image distortion model used in the above studies appeared to have only limited validity, the use of more abstract polynomial approximation methods for the description of image distortion was considered. The typically numerical problems encountered here have resulted in a separate investigation into the usefulness of bilinear least squares approximation, where substantial gains in computation time, memory requirements, and numerical stability may be achieved through the application of orthogonal transformations and separation of variables. Some theoretical and practical results are described in the last part of Chapter 2 (2.1.2 and 2.2). It would appear that such procedures may be generalized to 3-dimensional calibration of a complete camera configuration.

During the last stage of the present research these measurement devices and computational facilities were applied for the investigation of control processes of (fast) arm movements. With respect to fast motor actions especially, a distinction is usually made in the psychological and physiological litterature between decision and execution of (elementary) movements. It is investigated in Chapter 3 to what extent fast arm movements in one dimension may be readjusted if information on the distance to be covered becomes available only through feedback of the actual movement. The results suggested that here also occurs a ballistic, "distance-covering" phase which is followed by a phase of limited adjustment; in addition, differential effects were found for gradual and instantaneous target information presentation. In future research, control processes at a higher dimensionality will be investigated.

In Chapter 3.2 some preliminary results are presented with regard to the relation between locomotive behaviour and posture preservation, by means of correlation methods which relate various orthogonal movement components with each other.

Finally, in Chapter 4 a prospect and retrospect are presented, both as regards technology and as regards various application in psychology and related disciplines.

1977-4-10

Jucundi acti labores (sunt)

Cicero, De finibus bonorum et malorum, 2 32.105

4.3

Samenvatting

Dit onderzoek heeft betrekking op het meten en regelen van menselijke bewegingen, en valt in een meettechnisch en een gedragsmatig georienteerd deel uiteen.

Het onderzoek van de menselijke motoriek brengt ten gevolge van de grote variatie in mogelijke bewegingspatronen grote dataverwerkingsproblemen met zich mee. Een systeem dat volautomatische data-acquisitie en -reductie toelaat verdient dan ook de voorkeur boven meer traditionele benaderingen die veelal op - ten minste gedeeltelijke - manuele digitalisering van foto- of cinematografische gegevens berusten.

Een vergelijkend onderzoek naar enkele optoelectronische meetprocedures resulteerde in een experimenteel meetsysteem dat gebruik maakt van het *laterale fotoeffect* voor positiebepaling van kleine, op het lichaam aangebrachte lichtbronnen (1.2). In afwachting van een verbeterde, commerciele versie die in onafhankelijk onderzoek werd ontwikkeld aan de Chalmers Technische Hogeschool te Gothenburg, Zweden, werd een theoretisch onderzoek gedaan naar de eigenschappen van de in beide systemen toegepaste positiegevoelige omzetters (1.3).

Dit verbeterde, commerciele systeem liet door toepassing van twee camera's het meten toe van 3-dimensionale lichtpuntbanen, en een deel van het proefschrift is gewijd aan de ijk- en meetaspecten hiervan Hiertoe werd gebruik gemaakt van *fotogrammetrische* principes ter bepaling van systeem parameters zoals cameraposities en -houding, effectieve beeldafstand en beeldvervorming, door observatie van een verzameling ijkpunten die op een geschikt gekozen wijze over de waarnemingsruimte zijn verdeeld (Hoofdstuk 2).

Bij bewegingsonderzoek bestaat veelal behoefte aan simultane schatting van bewegingsbanen en afgeleiden naar de tijd, het gebruik van Kalman-filtermethoden werd besproken (2 1 2) en met een voorbeeld van 3-dimensionale gangreconstructie geillustreerd.

Omdat het hierboven gehanteerde model voor beeldvervormingscorrectie slechts beperkt bleek te functioneren, werd een onderzoek gedaan naar meer abstracte polynoombenadering voor de beschrijving van beeldvervorming. De typisch numerieke problemen die zich hierbij voordoen resulteerden in een afzonderlijk onderzoek naar de bruikbaarheid van bilineaire kleinste kwadraten benadering, waarbij door toepassing van orthogonale transformaties en scheiding van variabelen een grote winst in rekentijd, vereiste geheugenruimte en numerieke stabiliteit kan worden verkregen. Enkele theoretische en practische resultaten worden in het laatste deel van Hoofdstuk 2 beschreven (2.1.2 en 2.2). Het laat zich aanzien dat dergelijke procedures tot 3-dimensionale ijking van een complete camera opstelling kunnen worden gegeneraliseerd.

In de laatste fase van het promotie-onderzoek zijn deze meet- en rekentechnische hulpmiddelen ingezet voor de studie van besturingsprocessen van (snelle) armbewegingen. Met name bij snelle motorische handelingen wordt in de psychologische en fysiologische litteratuur een onderscheid gemaakt tussen beslissing en uitvoering van (elementaire) bewegingen. In Hoofdstuk 3.1 wordt nagegaan in hoeverre snelle armbewegingen in één dimensie kunnen worden bijgestuurd indien informatie over de af te leggen afstand eerst door terugkoppeling van de beweging zelve beschikbaar komt. De resultaten suggereerden dat ook hier sprake is van een ballistische, "distance-covering" fase gevolgd door een fase van beperkte bijsturing, waarbij differentiële effecten optraden voor geleidelijke en plotselinge presentatie van doelinformatie. In toekomstig onderzoek zullen stuurprocessen van hogere dimensionaliteit worden bezien.

In Hoofdstuk 3.2 worden enkele voorlopige resultaten gepresenteerd ten aanzien van de samenhang tussen ganggedrag en handhaving van het evenwicht, door middel van correlatiemethoden die bewegingen in verscheidene loodrechte richtingen aan elkaar relateren.

In Hoofdstuk 4 wordt tenslotte een terugblik en een vooruitzicht gegeven, zowel ten aanzien van de techniek als van diverse toepassingen binnen de Psychologie en haar randgebieden.

1977-4-10

Curriculum Vitae

Herman Jozef Woltring

1943-4-14	Geboren te Amsterdam
1961	Eindexamen Gymnasium β te Amsterdam
1967	Kandidaatsexamen Elektrotechniek, Technische Hogeschool Eindhoven
1969-1970	Studentassistentschap aan het Laboratorium voor Medische Fysica en Biofysica, Katholieke Universiteit Nijmegen, ten behoeve van een afstudeeronderzoek met betrekking tot het gebruik van continue ultrageluid voor detectie van vaatwandbewegingen
1970	Ingenieursexamen Elektrotechniek, Technische Hogeschool Eindhoven (Vakgroep Meet- en Regeltechniek)
1970-1971	Werkzaam in de vakgroep Bedrijfspsychologie, Katholieke Universiteit Nijmegen. Assistentie verricht in Ergonomisch onderzoek
1971-1972	Studie Experimentele Psychologie, Universiteit van Sussex, Engeland (ZWO/Royal Society European Science Exchange Fellowship)
1972	Master of Science examen Experimentele Psychologie, Universiteit van Sussex, Engeland
1972-1977	Werkzaam in de vakgroep Psychologie van Arbeid en Organisatie, Katholieke Universiteit Nijmegen. Onderzoek verricht naar instrumentatie voor bewegingsmeting bij de mens, naar verdere signaalbewerkingsmethoden hiervoor, en naar besturingsstrategieën van eenvoudige, snelle armbewegingen. Voor 1/3 deel gedetacheerd bij de Rekenmachinegroep van het Psychologisch Laboratorium ter assistentie voor experimenteel psychologisch onderzoek.

STELLINGEN

I. Met name ten aanzien van sociaal-psychologisch onderzoek - waarbij het aanbrengen van markeerpunten op proefpersonen op bezwaren stuit - is stelling VI van Ex, vanuit een practisch oogpunt bezien, nog steeds van toepassing.

De wetenschappelijke bestudering van de informatieve functie welke menselijke bewegingen, -handelwijzen en -gebaren bezitten, wordt aanzienlijk belemmerd door het gemis aan een preciese registratie-techniek. Slechts een apparatuur, welke in staat stelt deze bewegingen exact meetbaar te registreren - naar de vorm, de lengte en de richting van de afgelegde weg, naar de snelheid, versnelling en vertraging, en naar de afstand en de hoogte ten opzichte van de beweger - zou de belemmering kunnen opheffen. De constructie van zo'n apparatuur is voor de ontwikkeling van de psychologie, en in het bijzonder van de sociale psychologie, van groot belang.

J. Ex: Situatie-Analyse en Sociaal-Psychologisch Experiment. Proefschrift Nijmegen 1957.

II. Toepassing van 3-dimensionale reconstructiemethoden uit de analytische fotogrammetrie staat camera-opstellingen toe die meer door experimentele dan door meettechnische overwegingen worden beperkt. Dit is een aanmerkelijke verbetering ten opzichte van traditionele methoden in de biomechanica, waarbij camera's loodrecht ten opzichte van elkaar worden geacht of wier optische assen worden verondersteld elkaar te snijden.

III. De naar verhouding zeldzame, directe toepassing van orthogonale triangularisatie-methoden bij het oplossen van overgedetermineerde, lineaire vergelijkingsstelsels wekt verbazing indien men bedenkt dat de numeriek veelal slechter geconditioneerde oplossing via normaalvergelijkingen en Choleski-decompositie in feite eveneens van orthogonale triangularisatie gebruik maakt.

IV. Er bestaan geen biologische "open-loop" bewegingen in de strikte zin des woords.

V. Op zijn minst vanuit taalkundig oogpunt verdienen Gibson's definities van proprioceptie en kinesthese de voorkeur boven de alternatieve, door Hopkins aangegeven definities. Evenwel is er dan geen reden de kinesthese tot skeletale bewegingen te beperken.

J.J. Gibson: The Senses Considered as Perceptual Systems. London, Allen & Unwin 1968, p. 37.

B. Hopkins: Proprioception and/or Kinesthesis. Perceptual and Motor Skills *34*(1972) 431-435.

VI. Gezien het multidisciplinaire karakter van de Ergonomie verdient het aanbeveling deze niet bij "Bedrijfskunde" in de Technische Wetenschappen of "Arbeid en Organisatie" in de Sociologie of Psychologie onder te brengen, doch haar te laten aansluiten op een basisopleiding "Menswetenschappen".

Gegeven de medische, biologische, fysiologische, psychologische, linguïstische, fysische en mathematische aspecten van de psychonomie lijkt het wenselijk dat een nieuwe basisopleiding gecreëerd wordt in de *menswetenschappen*, van waaruit de specialisatie in de richting van hetzij de psychonomie, de humane biologie, de fysiologie of de humane biofysica (medische fysica) op natuurlijker wijze verloopt dan thans het geval is. Een van de voordelen van een dergelijke basisopleiding zou zijn, dat zij herkenbaar is voor studenten met de aanleg en belangstelling om zich in deze richting te ontwikkelen, waardoor een homogener studentenbestand zou worden aangetrokken dan thans met name bij de psychologie het geval is.

J.A. Michon, E.G.J. Eijkman en L.F.W. de Klerk: Handboek der Psychonomie. Van Loghum Slaterus, Deventer 1976, p. 624.

VII. Het verdient aanbeveling te onderzoeken of de vereisten voor het rijexamen dienen te worden uitgebreid met het slagen voor een cursus "Pech Onderweg", teneinde een minimum aan technisch inzicht bij gebruikers van potentieel gevaarlijke vervoersmiddelen bij te brengen. Eenzelfde eis ware te overwegen voor een E.H.B.O.-cursus.

VIII. De formele evaluatie van academische dissertaties dient voorbehouden te blijven aan de academie, en niet door politieke instellingen te worden overgenomen.

IX. In een tijd dat wetenschappelijk onderzoek sterk - en soms reeds bij voorbaat - op zijn maatschappelijke relevantie wordt getoetst, kunnen instellingen zoals de Stichting voor Zuiver Wetenschappelijk Onderzoek (Z.W.O.) de kwaliteit en coördinatie van onderzoek, zeker op langere termijn bezien, stimuleren en bewaken.

X. De Weert tekent in zijn stelling IX terecht bezwaar aan tegen het taalvervuilende gebruik van engelse of quasi-engelse woorden voor in goed nederlands te benoemen zaken. Voor wetenschappelijke begrippen dient evenwel gesteld te worden dat verregaande vernederlandsing van in goed engels-grieks-latijn te benoemen zaken een vorm van wetenschapsvervuiling is, en strijdig met de oorspronkelijke betekenis van *universitas* ofwel gemeenschap van volkeren.

Ch.M.M. de Weert: Binocular Colour and Brightness Combinations - an Experimental Study. Proefschrift, Nijmegen 1976.

XI. Het lezen van Asterix moge wellicht nationalistische gevoelens accentueren en geen wezenlijke bijdrage tot de eenwording van Europa leveren, zoals Van der Straaten heeft gesteld, hier staat tegenover dat de eenwording en groei van het onderzoek van menselijke bewegingen door dit lezen lijkt te worden gestimuleerd aangezien ook Dijkstra de gallische held naar voren haalt.

Stelling VIII, J.H.M. van der Straaten: De Activiteit van de Spieren rond het Kniegewricht tijdens het lopen op een Trottoir Roulant - een electromyografische studie. Proefschrift, Rotterdam 1972.

S. Dijkstra Evaluation and Simulation of Fast Arm Movements. Proefschrift, Utrecht 1973.

Nijmegen, 13 mei 1977 H.J. Woltring